PTO/PCT Rec'd 18 JAN 2002
PCT/US00/19774
10/031345

INDUCING CELLULAR IMMUNE RESPONSES TO HEPATITIS C VIRUS USING PEPTIDE AND NUCLEIC ACID COMPOSITIONS

FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

This invention was funded, in part, by the United States government under grants with the National Institutes of Health. The U.S. government has certain rights in this invention.

Jun. 21

INDEX

I. BACKGROUND OF THE INVENTION

Hepatitis C virus (HCV) infection is a global human health problem with approximately 150,000 new reported cases each year in the U.S. alone. HCV is a single stranded RNA virus, and is the etiological agent identified in most cases of non-A, non-B post-transfusion and post-transplant hepatitis, and is a common cause of acute sporadic hepatitis (Choo *et al.*, *Science* 244:359, 1989; Kuo *et al.*, *Science* 244:362, 1989; and Alter *et al.*, in: *Current Perspective in Hepatology*, p. 83, 1989). It is estimated that more than 50% of patients infected with HCV become chronically infected and, of those, 20% develop cirrhosis of the liver within 20 years (Davis *et al.*, *New Engl. J. Med.* 321:1501,

1989; Alter *et al.*, in: *Current Perspective in Hepatology*, p. 83, 1989; Alter *et al.*, *New Engl. J. Med.* 327:1899, 1992; and Dienstag, J. L. *Gastroenterology* 85:430, 1983). Moreover, the only therapy available for treatment of HCV infection is interferon-α. Most patients are unresponsive, however, and among the responders, there is a high recurrence rate within 6-12 months of cessation of treatment (Liang *et al.*, *J. Med. Virol.* 40:69, 1993). Ribaviron, a guanosine analog with a broad spectrum activity against many RNA and DNA viruses, has been shown in clinical trials to be effective against chronic HCV infection when used in combination with interferon- α (*see, e.g.*, Poynard *et al.*, *Lancet* 352:1426-1432, 1998; Reichard *et al.*, *Lancet* 351:83-87, 1998) However, the response rate is still well below 50%.

Virus-specific, human leukocyte antigen (HLA) class I-restricted cytotoxic T lymphocytes (CTL) are known to play a major role in the prevention and clearance of virus infections *in vivo* (Oldstone *et al.*, *Nature* 321:239, 1989; Jamieson *et al.*, *J. Virol.* 61:3930, 1987; Yap *et al*, *Nature* 273:238, 1978; Lukacher *et al.*, *J. Exp. Med.* 160:814, 1994; McMichael *et al.*, *N. Engl. J. Med.* 309:13, 1983; Sethi *et al.*, *J. Gen. Virol.* 64:443, 1983; Watari *et al.*, *J. Exp. Med.* 165:459, 1987; Yasukawa *et al.*, *J. Immunol.* 143:2051, 1989; Tigges *et al.*, *J. Virol.* 66:1622, 1993; Reddenhase *et al.*, *J. Virol.* 55:263, 1985; Quinnan *et al.*, *N. Engl. J. Med.* 307:6, 1982). HLA class I molecules are expressed on the surface of almost all nucleated cells. Following intracellular processing of antigens, epitopes from the antigens are presented as a complex with the HLA class I molecules on the surface of such cells. CTL recognize the peptide-HLA class I complex, which then results in the destruction of the cell bearing the HLA-peptide complex directly by the CTL and/or via the activation of non-destructive mechanisms *e.g.*, the production of interferon, that inhibit viral replication.

In view of the heterogeneous immune response observed with HCV infection, induction of a multi-specific cellular immune response directed simultaneously against multiple HCV epitopes appears to be important for the development of an efficacious vaccine against HCV. There is a need, however, to establish vaccine embodiments that elicit immune responses that correspond to responses seen in patients that clear HCV infection.

The information provided in this section is intended to disclose the presently understood state of the art as of the filing date of the present application. Information is included in this section which was generated subsequent to the priority date of this

application. Accordingly, information in this section is not intended, in any way, to delineate the priority date for the invention.

II. SUMMARY OF THE INVENTION

This invention applies our knowledge of the mechanisms by which antigen is recognized by T cells, for example, to develop epitope-based vaccines directed towards HCV. More specifically, this application communicates our discovery of specific epitope pharmaceutical compositions and methods of use in the prevention and treatment of HCV infection.

Upon development of appropriate technology, the use of epitope-based vaccines has several advantages over current vaccines, particularly when compared to the use of whole antigens in vaccine compositions. There is evidence that the immune response to whole antigens is directed largely toward variable regions of the antigen, allowing for immune escape due to mutations. The epitopes for inclusion in an epitope-based vaccine are selected from conserved regions of viral or tumor-associated antigens, which thereby reduces the likelihood of escape mutants. Furthermore, immunosuppressive epitopes that may be present in whole antigens can be avoided with the use of epitope-based vaccines.

An additional advantage of an epitope-based vaccine approach is the ability to combine selected epitopes (CTL and HTL), and further, to modify the composition of the epitopes, achieving, for example, enhanced immunogenicity. Accordingly, the immune response can be modulated, as appropriate, for the target disease. Similar engineering of the response is not possible with traditional approaches.

Another major benefit of epitope-based immune-stimulating vaccines is their safety. The possible pathological side effects caused by infectious agents or whole protein antigens, which might have their own intrinsic biological activity, is eliminated.

An epitope-based vaccine also provides the ability to direct and focus an immune response to multiple selected antigens from the same pathogen. Thus, patient-by-patient variability in the immune response to a particular pathogen may be alleviated by inclusion of epitopes from multiple antigens from that pathogen in a vaccine composition. A "pathogen" may be an infectious agent or a tumor associated molecule.

One of the most formidable obstacles to the development of broadly efficacious epitope-based immunotherapeutics, however, has been the extreme polymorphism of HLA molecules. To date, effective non-genetically biased coverage of a population has been a task of considerable complexity; such coverage has required that epitopes be used

that are specific for HLA molecules corresponding to each individual HLA allele, therefore, impractically large numbers of epitopes would have to be used in order to cover ethnically diverse populations. Thus, there has existed a need for peptide epitopes that are bound by multiple HLA antigen molecules for use in epitope-based vaccines. The greater the number of HLA antigen molecules bound, the greater the breadth of population coverage by the vaccine.

Furthermore, as described herein in greater detail, a need has existed to modulate peptide binding properties, for example, so that peptides that are able to bind to multiple HLA antigens do so with an affinity that will stimulate an immune response. Identification of epitopes restricted by more than one HLA allele at an affinity that correlates with immunogenicity is important to provide thorough population coverage, and to allow the elicitation of responses of sufficient vigor to prevent or clear an infection in a diverse segment of the population. Such a response can also target a broad array of epitopes. The technology disclosed herein provides for such favored immune responses.

In a preferred embodiment, epitopes for inclusion in vaccine compositions of the invention are selected by a process whereby protein sequences of known antigens are evaluated for the presence of motif or supermotif-bearing epitopes. Peptides corresponding to a motif- or supermotif-bearing epitope are then synthesized and tested for the ability to bind to the HLA molecule that recognizes the selected motif. Those peptides that bind at an intermediate or high affinity *i.e.*, an IC_{50} (or a K_D value) of 500 nM or less for HLA class I molecules or 1000 nM or less for HLA class II molecules, are further evaluated for their ability to induce a CTL or HTL response. Immunogenic peptide epitopes are selected for inclusion in vaccine compositions.

Supermotif-bearing peptides may additionally be tested for the ability to bind to multiple alleles within the HLA supertype family. Moreover, peptide epitopes may be analogued to modify binding affinity and/or the ability to bind to multiple alleles within an HLA supertype.

The invention also includes an embodiment comprising a method for monitoring or evaluating an immune response to HCV in a patient having a known HLA-type, the method comprising incubating a T lymphocyte sample from the patient with a peptide composition comprising an HCV epitope consisting essentially of an amino acid sequence described in Tables VII to Table XX or Table XXII which binds the product of at least one HLA allele present in said patient, and detecting for the presence of a T lymphocyte

that binds to the peptide. A CTL peptide epitope may, for example, comprise a tetrameric complex.

An alternative modality for defining the peptide epitopes in accordance with the invention is to recite the physical properties, such as length; primary structure; or charge, which are correlated with binding to a particular allele-specific HLA molecule or group of allele-specific HLA molecules. A further modality for defining peptide epitopes is to recite the physical properties of an HLA binding pocket, or properties shared by several allele-specific HLA binding pockets (*e.g.* pocket configuration and charge distribution) and reciting that the peptide epitope fits and binds to said pocket or pockets.

As will be apparent from the discussion below, other methods and embodiments are also contemplated. Further, novel synthetic peptides produced by any of the methods described herein are also part of the invention.

III. BRIEF DESCRIPTION OF THE FIGURES

Figure 1: Figure 1 provides a graph of total frequency of genotypes as a function of the number of HCV candidate epitopes bound by HLA-A and B molecules, in an average population.

Figure 2: Figure 2 illustrates the position of peptide epitopes in an experimental model minigene construct.

IV. DETAILED DESCRIPTION OF THE INVENTION

The peptide epitopes and corresponding nucleic acid compositions of the present invention are useful for stimulating an immune response to HCV by stimulating the production of CTL or HTL responses. The peptide epitopes, which are derived directly or indirectly from native HCV amino acid sequences, are able to bind to HLA molecules and stimulate an immune response to HCV. The complete polyprotein sequence from HCV and its variants can be obtained from Genbank. Peptide epitopes and analogs thereof can also be readily determined from sequence information that may subsequently be discovered for heretofore unknown variants of HCV, as will be clear from the disclosure provided below.

The peptide epitopes of the invention have been identified in a number of ways, as will be discussed below. Also discussed in greater detail is that analog peptides have been derived and the binding activity for HLA molecules modulated by modifying specific amino acid residues to create peptide analogs exhibiting altered immunogenicity.

Further, the present invention provides compositions and combinations of compositions that enable epitope-based vaccines that are capable of interacting with HLA molecules encoded by various genetic alleles to provide broader population coverage than prior vaccines.

IV.A. Definitions

The invention can be better understood with reference to the following definitions, which are listed alphabetically:

A “computer” or “computer system” generally includes: a processor; at least one information storage/retrieval apparatus such as, for example, a hard drive, a disk drive or a tape drive; at least one input apparatus such as, for example, a keyboard, a mouse, a touch screen, or a microphone; and display structure. Additionally, the computer may include a communication channel in communication with a network. Such a computer may include more or less than what is listed above.

“Cross-reactive binding” indicates that a peptide is bound by more than one HLA molecule; a synonym is degenerate binding.

A "cryptic epitope" elicits a response by immunization with an isolated peptide, but the response is not cross-reactive *in vitro* when intact whole protein which comprises the epitope is used as an antigen.

A "dominant epitope" is an epitope that induces an immune response upon immunization with a whole native antigen (see, *e.g.*, Sercarz, *et al.*, *Annu. Rev. Immunol.* 11:729-766, 1993). Such a response is cross-reactive *in vitro* with an isolated peptide epitope.

With regard to a particular amino acid sequence, an “epitope” is a set of amino acid residues which is involved in recognition by a particular immunoglobulin, or in the context of T cells, those residues necessary for recognition by T cell receptor proteins and/or Major Histocompatibility Complex (MHC) receptors. In an immune system setting, *in vivo* or *in vitro*, an epitope is the collective features of a molecule, such as primary, secondary and tertiary peptide structure, and charge, that together form a site recognized by an immunoglobulin, T cell receptor or HLA molecule. Throughout this disclosure epitope and peptide are often used interchangeably.

It is to be appreciated that protein or peptide molecules that comprise an epitope of the invention as well as additional amino acid(s) are still within the bounds of the invention. In certain embodiments, there is a limitation on the length of a peptide of the

invention which is not otherwise a construct. An embodiment that is length-limited occurs when the protein/peptide comprising an epitope of the invention comprises a region (i.e., a contiguous series of amino acids) having 100% identity with a native sequence. In order to avoid the definition of epitope from reading, *e.g.*, on whole natural molecules, there is a limitation on the length of any region that has 100% identity with a native peptide sequence. Thus, for a peptide comprising an epitope of the invention and a region with 100% identity with a native peptide sequence (and is not otherwise a construct), the region with 100% identity to a native sequence generally has a length of: less than or equal to 600 amino acids, often less than or equal to 500 amino acids, often less than or equal to 400 amino acids, often less than or equal to 250 amino acids, often less than or equal to 100 amino acids, often less than or equal to 85 amino acids, often less than or equal to 75 amino acids, often less than or equal to 65 amino acids, and often less than or equal to 50 amino acids. In certain embodiments, an "epitope" of the invention is comprised by a peptide having a region with less than 51 amino acids that has 100% identity to a native peptide sequence, in any increment of (49, 48, 47, 46, 45, 44, 43, 42, 41, 40, 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5) down to 5 amino acids.

Accordingly, peptide or protein sequences longer than 600 amino acids are within the scope of the invention, so long as they do not comprise any contiguous sequence of more than 600 amino acids that have 100% identity with a native peptide sequence, if they are not otherwise a construct. For any peptide that has five contiguous residues or less that correspond to a native sequence, there is no limitation on the maximal length of that peptide in order to fall within the scope of the invention. It is presently preferred that a CTL epitope be less than 600 residues long in any increment down to eight amino acid residues.

"Human Leukocyte Antigen" or "HLA" is a human class I or class II Major Histocompatibility Complex (MHC) protein (*see*, *e.g.*, Stites, *et al.*, IMMUNOLOGY, 8TH ED., Lange Publishing, Los Altos, CA (1994).

An "HLA supertype or family", as used herein, describes sets of HLA molecules grouped on the basis of shared peptide-binding specificities. HLA class I molecules that share somewhat similar binding affinity for peptides bearing certain amino acid motifs are grouped into HLA supertypes. The terms HLA superfamily, HLA supertype family, HLA family, and HLA xx-like supertype molecules (where xx denotes a particular HLA type), are synonyms.

Throughout this disclosure, results are expressed in terms of "IC_{50}'s." IC_{50} is the concentration of peptide in a binding assay at which 50% inhibition of binding of a reference peptide is observed. Given the conditions in which the assays are run (*i.e.*, limiting HLA proteins and labeled peptide concentrations), these values approximate K_D values. Assays for determining binding are described in detail, *e.g.*, in PCT publications WO 94/20127 and WO 94/03205. It should be noted that IC_{50} values can change, often dramatically, if the assay conditions are varied, and depending on the particular reagents used (*e.g.*, HLA preparation, *etc.*). For example, excessive concentrations of HLA molecules will increase the apparent measured IC_{50} of a given ligand.

Alternatively, binding is expressed relative to a reference peptide. Although as a particular assay becomes more, or less, sensitive, the IC_{50}'s of the peptides tested may change somewhat, the binding relative to the reference peptide will not significantly change. For example, in an assay run under conditions such that the IC_{50} of the reference peptide increases 10-fold, the IC_{50} values of the test peptides will also shift approximately 10-fold. Therefore, to avoid ambiguities, the assessment of whether a peptide is a good, intermediate, weak, or negative binder is generally based on its IC_{50}, relative to the IC_{50} of a standard peptide.

Binding may also be determined using other assay systems including those using: live cells (*e.g.*, Ceppellini *et al.*, *Nature* 339:392, 1989; Christnick *et al.*, *Nature* 352:67, 1991; Busch *et al.*, *Int. Immunol.* 2:443, 19990; Hill *et al.*, *J. Immunol.* 147:189, 1991; del Guercio *et al.*, *J. Immunol.* 154:685, 1995), cell free systems using detergent lysates (*e.g.*, Cerundolo *et al.*, *J. Immunol.* 21:2069, 1991), immobilized purified MHC (*e.g.*, Hill *et al.*, *J. Immunol.* 152, 2890, 1994; Marshall *et al.*, *J. Immunol.* 152:4946, 1994), ELISA systems (*e.g.*, Reay *et al.*, *EMBO J.* 11:2829, 1992), surface plasmon resonance (*e.g.*, Khilko *et al.*, *J. Biol. Chem.* 268:15425, 1993); high flux soluble phase assays (Hammer *et al.*, *J. Exp. Med.* 180:2353, 1994), and measurement of class I MHC stabilization or assembly (*e.g.*, Ljunggren *et al.*, *Nature* 346:476, 1990; Schumacher *et al.*, *Cell* 62:563, 1990; Townsend *et al.*, *Cell* 62:285, 1990; Parker *et al.*, *J. Immunol.* 149:1896, 1992).

As used herein, "high affinity" with respect to HLA class I molecules is defined as binding with an IC_{50}, or K_D value, of 50 nM or less; "intermediate affinity" is binding with an IC_{50} or K_D value of between about 50 and about 500 nM. "High affinity" with respect to binding to HLA class II molecules is defined as binding with an IC_{50} or K_D value of 100 nM or less; "intermediate affinity" is binding with an IC_{50} or K_D value of between about 100 and about 1000 nM.

The terms "identical" or percent "identity," in the context of two or more peptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues that are the same, when compared and aligned for maximum correspondence over a comparison window, as measured using a sequence comparison algorithm or by manual alignment and visual inspection.

An "immunogenic peptide" or "peptide epitope" is a peptide that comprises an allele-specific motif or supermotif such that the peptide will bind an HLA molecule and induce a CTL and/or HTL response. Thus, immunogenic peptides of the invention are capable of binding to an appropriate HLA molecule and thereafter inducing an HLA-restricted cytotoxic or helper T cell response to the antigen from which the immunogenic peptide is derived.

The phrases "isolated" or "biologically pure" refer to material which is substantially or essentially free from components which normally accompany the material as it is found in its native state. Thus, isolated peptides in accordance with the invention preferably do not contain materials normally associated with the peptides in their *in situ* environment. An "isolated" epitope refers to an epitope that does not include the whole sequence of the antigen or polypeptide from which the epitope was derived. Typically the "isolated" epitope does not have attached thereto additional amino acids that result in a sequence that has 100% identity with a native sequence. The native sequence can be a sequence such as a tumor-associated antigen from which the epitope is derived.

"Major Histocompatibility Complex" or "MHC" is a cluster of genes that plays a role in control of the cellular interactions responsible for physiologic immune responses. In humans, the MHC complex is also known as the HLA complex. For a detailed description of the MHC and HLA complexes, see, Paul, FUNDAMENTAL IMMUNOLOGY, 3RD ED., Raven Press, New York, 1993.

The term "motif" refers to the pattern of residues in a peptide of defined length, usually a peptide of from about 8 to about 13 amino acids for a class I HLA motif and from about 6 to about 25 amino acids for a class II HLA motif, which is recognized by a particular HLA molecule. Peptide motifs are typically different for each protein encoded by each human HLA allele and differ in the pattern of the primary and secondary anchor residues.

A "negative binding residue" is an amino acid which, if present at certain positions (typically not primary anchor positions) in a peptide epitope, results in decreased binding affinity of the peptide for the peptide's corresponding HLA molecule.

A “non-native” sequence or “construct” refers to a sequence that is not found in in nature (“non-naturally occurring”). Such sequences include, *e.g.*, peptides that are lipidated or otherwise modifed and polyepitopic compositions that contain epitopes that are non contiguous in a native protein sequence.

The term "peptide" is used interchangeably with "oligopeptide" in the present specification to designate a series of residues, typically L-amino acids, connected one to the other, typically by peptide bonds between the α-amino and carboxyl groups of adjacent amino acids. The preferred CTL-inducing peptides of the invention are 13 residues or less in length and usually consist of between about 8 and about 11 residues, preferably 9 or 10 residues. The preferred HTL-inducing oligopeptides are less than about 50 residues in length and usually consist of between about 6 and about 30 residues, more usually between about 12 and 25, and often between about 15 and 20 residues.

“Pharmaceutically acceptable” refers to a generally non-toxic, inert, and/or physiologically compatible composition.

A “pharmaceutical excipient” comprises a material such as an adjuvant, a carrier, pH-adjusting and buffering agents, tonicity adjusting agents, wetting agents, preservative, and the like.

A "primary anchor residue" is an amino acid at a specific position along a peptide sequence which is understood to provide a contact point between the immunogenic peptide and the HLA molecule. One to three, usually two, primary anchor residues within a peptide of defined length generally defines a “motif” for an immunogenic peptide. These residues are understood to fit in close contact with peptide binding grooves of an HLA molecule, with their side chains buried in specific pockets of the binding grooves themselves. In one embodiment, the primary anchor residues are located at position 2 (from the amino terminal position) and at the carboxyl terminal position of a 9-residue peptide epitope in accordance with the invention. The primary anchor positions for each motif and supermotif are set forth in Table 1. For example, analog peptides can be created by altering the presence or absence of particular residues in these primary anchor positions. Such analogs are used to modulate the binding affinity of a peptide comprising a particular motif or supermotif.

“Promiscuous recognition” is where a distinct peptide is recognized by the same T cell clone in the context of various HLA molecules. Promiscuous recognition or binding is synonymous with cross-reactive binding.

A "protective immune response" or "therapeutic immune response" refers to a CTL and/or an HTL response to an antigen derived from an infectious agent or a tumor antigen, which prevents or at least partially arrests disease symptoms or progression. The immune response may also include an antibody response which has been facilitated by the stimulation of helper T cells.

The term "residue" refers to an amino acid or amino acid mimetic incorporated into an oligopeptide by an amide bond or amide bond mimetic.

A "secondary anchor residue" is an amino acid at a position other than a primary anchor position in a peptide which may influence peptide binding. A secondary anchor residue occurs at a significantly higher frequency amongst bound peptides than would be expected by random distribution of amino acids at one position. The secondary anchor residues are said to occur at "secondary anchor positions." A secondary anchor residue can be identified as a residue which is present at a higher frequency among high or intermediate affinity binding peptides, or a residue otherwise associated with high or intermediate affinity binding. For example, analog peptides can be created by altering the presence or absence of particular residues in these secondary anchor positions. Such analogs are used to finely modulate the binding affinity of a peptide comprising a particular motif or supermotif.

A "subdominant epitope" is an epitope which evokes little or no response upon immunization with whole antigens which comprise the epitope, but for which a response can be obtained by immunization with an isolated peptide, and this response (unlike the case of cryptic epitopes) is detected when whole protein is used to recall the response *in vitro* or *in vivo*.

A "supermotif" is a peptide binding specificity shared by HLA molecules encoded by two or more HLA alleles. Preferably, a supermotif-bearing peptide is recognized with high or intermediate affinity (as defined herein) by two or more HLA antigens.

"Synthetic peptide" refers to a peptide that is man-made using such methods as chemical synthesis or recombinant DNA technology.

As used herein, a "vaccine" is a composition that contains one or more peptides of the invention. There are numerous embodiments of vaccines in accordance with the invention, such as by a cocktail of one or more peptides; one or more epitopes of the invention comprised by a polyepitopic peptide; or nucleic acids that encode such peptides or polypeptides, e.g., a minigene that encodes a polyepitopic peptide. The "one or more peptides" can include, e.g., at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,

19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40 , 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 or more peptides of the invention. The peptides or polypeptides can optionally be modified, such as by lipidation, addition of targeting or other sequences. HLA class I-binding peptides of the invention can be admixed with, or linked to, HLA class II-binding peptides, to facilitate activation of both cytotoxic T lymphocytes and helper T lymphocytes. Vaccines can also comprise peptide-pulsed antigen presenting cells, e.g., dendritic cells.

The nomenclature used to describe peptide compounds follows the conventional practice wherein the amino group is presented to the left (the N-terminus) and the carboxyl group to the right (the C-terminus) of each amino acid residue. When amino acid residue positions are referred to in a peptide epitope they are numbered in an amino to carboxyl direction with position one being the position closest to the amino terminal end of the epitope, or the peptide or protein of which it may be a part. In the formulae representing selected specific embodiments of the present invention, the amino- and carboxyl-terminal groups, although not specifically shown, are in the form they would assume at physiologic pH values, unless otherwise specified. In the amino acid structure formulae, each residue is generally represented by standard three letter or single letter designations. The L-form of an amino acid residue is represented by a capital single letter or a capital first letter of a three-letter symbol, and the D-form for those amino acids having D-forms is represented by a lower case single letter or a lower case three letter symbol. Glycine has no asymmetric carbon atom and is simply referred to as "Gly" or G. Symbols for the amino acids are shown below.

Single Letter Symbol	Three Letter Symbol	Amino Acids
A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic Acid
E	Glu	Glutamic Acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

IV.B. Stimulation of CTL and HTL responses

The mechanism by which T cells recognize antigens has been delineated during the past ten years. Based on our understanding of the immune system we have developed efficacious peptide epitope vaccine compositions that can induce a therapeutic or prophylactic immune response to HCV in a broad population. For an understanding of the value and efficacy of the claimed compositions, a brief review of immunology-related technology is provided.

A complex of an HLA molecule and a peptidic antigen acts as the ligand recognized by HLA-restricted T cells (Buus, S. *et al.*, *Cell* 47:1071, 1986; Babbitt, B. P. *et al.*, *Nature* 317:359, 1985; Townsend, A. and Bodmer, H., *Annu. Rev. Immunol.* 7:601,

1989; Germain, R. N., *Annu. Rev. Immunol.* 11:403, 1993). Through the study of single amino acid substituted antigen analogs and the sequencing of endogenously bound, naturally processed peptides, critical residues that correspond to motifs required for specific binding to HLA antigen molecules have been identified and are described herein and are set forth in Tables I, II, and III (*see* also, *e.g.*, Southwood, *et al.*, *J. Immunol.* 160:3363, 1998; Rammensee, *et al.*, *Immunogenetics* 41:178, 1995; Rammensee *et al.*, SYFPEITHI, access via web at : http://134.2.96.221/scripts.hlaserver.dll/home.htm; Sette, A. and Sidney, J. *Curr. Opin. Immunol.* 10:478, 1998; Engelhard, V. H., *Curr. Opin. Immunol.* 6:13, 1994; Sette, A. and Grey, H. M., *Curr. Opin. Immunol.* 4:79, 1992; Sinigaglia, F. and Hammer, J. *Curr. Biol.* 6:52, 1994; Ruppert *et al.*, *Cell* 74:929-937, 1993; Kondo *et al.*, *J. Immunol.* 155:4307-4312, 1995; Sidney *et al.*, *J. Immunol.* 157:3480-3490, 1996; Sidney *et al.*, *Human Immunol.* 45:79-93, 1996; Sette, A. and Sidney, J. *Immunogenetics*, in press, 1999).

Furthermore, x-ray crystallographic analysis of HLA-peptide complexes has revealed pockets within the peptide binding cleft of HLA molecules which accommodate, in an allele-specific mode, residues borne by peptide ligands; these residues in turn determine the HLA binding capacity of the peptides in which they are present. (*See, e.g.*, Madden, D.R. *Annu. Rev. Immunol.* 13:587, 1995; Smith, *et al.*, *Immunity* 4:203, 1996; Fremont *et al.*, *Immunity* 8:305, 1998; Stern *et al.*, *Structure* 2:245, 1994; Jones, E.Y. *Curr. Opin. Immunol.* 9:75, 1997; Brown, J. H. *et al.*, *Nature* 364:33, 1993; Guo, H. C. *et al.*, *Proc. Natl. Acad. Sci. USA* 90:8053, 1993; Guo, H. C. *et al.*, *Nature* 360:364, 1992; Silver, M. L. *et al.*, *Nature* 360:367, 1992; Matsumura, M. *et al.*, *Science* 257:927, 1992; Madden *et al.*, *Cell* 70:1035, 1992; Fremont, D. H. *et al.*, *Science* 257:919, 1992; Saper, M. A. , Bjorkman, P. J. and Wiley, D. C., *J. Mol. Biol.* 219:277, 1991.)

Accordingly, the definition of class I and class II allele-specific HLA binding motifs, or class I or class II supermotifs allows identification of regions within a protein that have the potential of binding particular HLA antigen(s).

The present inventors have found that the correlation of binding affinity with immunogenicity, which is disclosed herein, is an important factor to be considered when evaluating candidate peptides. Thus, by a combination of motif searches and HLA-peptide binding assays, candidates for epitope-based vaccines have been identified. After determining their binding affinity, additional confirmatory work can be performed to select, amongst these vaccine candidates, epitopes with preferred characteristics in terms of population coverage, antigenicity, and immunogenicity.

Various strategies can be utilized to evaluate immunogenicity, including:

1) Evaluation of primary T cell cultures from normal individuals (*see, e.g.*, Wentworth, P. A. *et al., Mol. Immunol.* 32:603, 1995; Celis, E. *et al., Proc. Natl. Acad. Sci. USA* 91:2105, 1994; Tsai, V. *et al., J. Immunol.* 158:1796, 1997; Kawashima, I. *et al., Human Immunol.* 59:1, 1998); This procedure involves the stimulation of peripheral blood lymphocytes (PBL) from normal subjects with a test peptide in the presence of antigen presenting cells *in vitro* over a period of several weeks. T cells specific for the peptide become activated during this time and are detected using, *e.g.*, a ^{51}Cr-release assay involving peptide sensitized target cells.

2) Immunization of HLA transgenic mice (*see, e.g.*, Wentworth, P. A. *et al., J. Immunol.* 26:97, 1996; Wentworth, P. A. *et al., Int. Immunol.* 8:651, 1996; Alexander, J. *et al., J. Immunol.* 159:4753, 1997); In this method, peptides in incomplete Freund's adjuvant are administered subcutaneously to HLA transgenic mice. Several weeks following immunization, splenocytes are removed and cultured *in vitro* in the presence of test peptide for approximately one week. Peptide-specific T cells are detected using, *e.g.*, a ^{51}Cr-release assay involving peptide sensitized target cells and target cells expressing endogenously generated antigen.

3) Demonstration of recall T cell responses from immune individuals who have effectively been vaccinated, recovered from infection, and/or from chronically infected patients (*see, e.g.*, Rehermann, B. *et al., J. Exp. Med.* 181:1047, 1995; Doolan, D. L. *et al., Immunity* 7:97, 1997; Bertoni, R. *et al., J. Clin. Invest.* 100:503, 1997; Threlkeld, S. C. *et al., J. Immunol.* 159:1648, 1997; Diepolder, H. M. *et al., J. Virol.* 71:6011, 1997). In applying this strategy, recall responses are detected by culturing PBL from subjects that have been naturally exposed to the antigen, for instance through infection, and thus have generated an immune response "naturally", or from patients who were vaccinated against the infection. PBL from subjects are cultured *in vitro* for 1-2 weeks in the presence of test peptide plus antigen presenting cells (APC) to allow activation of "memory" T cells, as compared to "naive" T cells. At the end of the culture period, T cell activity is detected using assays for T cell activity including ^{51}Cr release involving peptide-sensitized targets, T cell proliferation, or lymphokine release.

The following describes the peptide epitopes and corresponding nucleic acids of the invention.

IV.C. Binding Affinity of Peptide Epitopes for HLA Molecules

The large degree of HLA polymorphism is an important factor to consider with the epitope-based approach to vaccine development. To address this factor, epitope selection including identification of peptides capable of binding at high or intermediate affinity to multiple HLA molecules is often utilized, most preferably these epitopes bind at high or intermediate affinity to two or more allele specific HLA molecules.

CTL-inducing peptides of interest for vaccine compositions preferably include those that have an IC_{50} or binding affinity value for class I HLA molecules of 500 nM or better (*i.e.*, the value is ≤ 500 nM). HTL-inducing peptides preferably include those that have an IC_{50} or binding affinity value for class II HLA molecules of 1000 nM or better, (*i.e.*, the value is ≤ 1,000 nM). For example, peptide binding is assessed by testing the capacity of a candidate peptide to bind to a purified HLA molecule *in vitro*. Peptides exhibiting high or intermediate affinity are then considered for further analysis. Selected peptides are tested on other members of the supertype family. In preferred embodiments, peptides that exhibit cross-reactive binding are then used in vaccines or in cellular screening analyses.

Higher HLA binding affinity is typically correlated with greater immunogenicity. Greater immunogenicity can be manifested in several different ways. Immunogenicity corresponds to whether an immune response is elicited at all, and to the vigor of any particular response, as well as to the extent of a population in which a response is elicited. For example, a peptide might elicit an immune response in a diverse array of the population, yet in no instance produce a vigorous response. In accordance with these principles, close to 90% of high binding peptides have been found to be immunogenic, as contrasted with about 50% of the peptides which bind with intermediate affinity. Moreover, higher binding affinity peptides leads to more vigorous immunogenic responses. As a result, less peptide is required to elicit a similar biological effect if a high affinity binding peptide is used. Thus, in preferred embodiments of the invention, high affinity binding epitopes are particularly useful.

The relationship between binding affinity for HLA class I molecules and immunogenicity of discrete peptide epitopes on bound antigens has been determined for the first time in the art by the present inventors. The correlation between binding affinity and immunogenicity was analyzed in two different experimental approaches (*see, e.g.*, Sette, *et al., J. Immunol.* 153:5586-5592, 1994). In the first approach, the

immunogenicity of potential epitopes ranging in HLA binding affinity over a 10,000-fold range was analyzed in HLA-A*0201 transgenic mice. In the second approach, the antigenicity of approximately 100 different hepatitis B virus (HBV)-derived potential epitopes, all carrying A*0201 binding motifs, was assessed by using PBL from acute hepatitis patients. Pursuant to these approaches, it was determined that an affinity threshold value of approximately 500 nM (preferably 50 nM or less) determines the capacity of a peptide epitope to elicit a CTL response. These data are true for class I binding affinity measurements for naturally processed peptides and for synthesized T cell epitopes. These data also indicate the important role of determinant selection in the shaping of T cell responses (*see, e.g.,* Schaeffer *et al. Proc. Natl. Acad. Sci. USA* 86:4649-4653, 1989).

An affinity threshold associated with immunogenicity in the context of HLA class II DR molecules has also been delineated (*see, e.g.*, Southwood *et al. J. Immunology* 160:3363-3373,1998). In order to define a biologically significant threshold of DR binding affinity, a database of the binding affinities of 32 DR-restricted epitopes for their restricting element (*i.e.*, the HLA molecule that binds the motif) was compiled. In approximately half of the cases (15 of 32 epitopes), DR restriction was associated with high binding affinities, *i.e.* binding affinity values of 100 nM or less. In the other half of the cases (16 of 32), DR restriction was associated with intermediate affinity (binding affinity values in the 100-1000 nM range). In only one of 32 cases was DR restriction associated with an IC_{50} of 1000 nM or greater. Thus, 1000 nM can be defined as an affinity threshold associated with immunogenicity in the context of DR molecules.

The binding affinity of peptides for HLA molecules can be determined as described in Example 1, below.

IV.D. Peptide Epitope Binding Motifs and Supermotifs

In the past few years evidence has accumulated to demonstrate that a large fraction of HLA class I and class II molecules can be classified into a relatively few supertypes, each characterized by largely overlapping peptide binding repertoires, and consensus structures of the main peptide binding pockets.

For HLA molecule pocket analyses, the residues comprising the B and F pockets of HLA class I molecules as described in crystallographic studies were analyzed (*see, e.g.*, Guo, H. C. *et al.*, *Nature* 360:364, 1992; Saper, M. A. , Bjorkman, P. J. and Wiley, D. C., *J. Mol. Biol.* 219:277, 1991; Madden, D. R., Garboczi, D. N. and Wiley, D. C.,

Cell 75:693, 1993; Parham, P., Adams, E. J., and Arnett, K. L., *Immunol. Rev.* 143:141, 1995). In these analyses, residues 9, 45, 63, 66, 67, 70, and 99 were considered to make up the B pocket; and the B pocket was deemed to determine the specificity for the amino acid residue in the second position of peptide ligands. Similarly, residues 77, 80, 81, and 116 were considered to determine the specificity of the F pocket; the F pocket was deemed to determine the specificity for the C-terminal residue of a peptide ligand bound by the HLA class I molecule.

Through the study of single amino acid substituted antigen analogs and the sequencing of endogenously bound, naturally processed peptides, critical residues required for allele-specific binding to HLA molecules have been identified. The presence of these residues correlates with binding affinity for HLA molecules. The identification of motifs and/or supermotifs that correlate with high and intermediate affinity binding is an important issue with respect to the identification of immunogenic peptide epitopes for the inclusion in a vaccine. Kast *et al.* (*J. Immunol.* 152:3904-3912, 1994) have shown that motif-bearing peptides account for 90% of the epitopes that bind to allele-specific HLA class I molecules. In this study all possible peptides of 9 amino acids in length and overlapping by eight amino acids (240 peptides), which cover the entire sequence of the E6 and E7 proteins of human papillomavirus type 16, were evaluated for binding to five allele-specific HLA molecules that are expressed at high frequency among different ethnic groups. This unbiased set of peptides allowed an evaluation of the predictive value of HLA class I motifs. From the set of 240 peptides, 22 peptides were identified that bound to an allele-specific HLA molecule with high or intermediate affinity. Of these 22 peptides, 20 (*i.e.* 91%) were motif-bearing. Thus, this study demonstrates the value of motifs for the identification of peptide epitopes for inclusion in a vaccine: application of motif-based identification techniques eliminates screening of 90% of the potential epitopes in a target antigen protein sequence.

Such peptide epitopes are identified in the Tables described below.

Peptides of the present invention may also comprise epitopes that bind to MHC class II DR molecules. A greater degree of heterogeneity in both size and binding frame position of the motif, relative to the N and C termini of the peptide, exists for class II peptide ligands. This increased heterogeneity of HLA class II peptide ligands is due to the structure of the binding groove of the HLA class II molecule which, unlike its class I counterpart, is open at both ends. Crystallographic analysis of HLA class II DRB*0101-peptide complexes showed that the major energy of binding is contributed by peptide

residues complexed with complementary pockets on the DRB*0101 molecules. An important anchor residue engages the deepest hydrophobic pocket (*see, e.g.*, Madden, D.R. *Ann. Rev. Immunol.* 13:587, 1995) and is referred to as position 1 (P1). P1 may represent the N-terminal residue of a class II binding peptide epitope, but more typically is flanked towards the N-terminus by one or more residues. Other studies have also pointed to an important role for the peptide residue in the 6th position towards the C-terminus, relative to P1, for binding to various DR molecules.

Thus, peptides of the present invention are identified by any one of several HLA-specific amino acid motifs (*see, e.g.*, Tables I-III). If the presence of the motif corresponds to the ability to bind several allele-specific HLA antigens, it is referred to as a supermotif. The HLA molecules that bind to peptides that possess a particular amino acid supermotif are collectively referred to as an HLA "supertype."

The peptide motifs and supermotifs described below, and summarized in Tables I-III, provide guidance for the identification and use of peptide epitopes in accordance with the invention.

Examples of peptide epitopes bearing a respective supermotif or motif are included in Tables as designated in the description of each motif or supermotif below. The Tables include a binding affinity ratio listing for some of the peptide epitopes. The ratio may be converted to IC_{50} by using the following formula: IC_{50} of the standard peptide/ratio = IC_{50} of the test peptide (*i.e.*, the peptide epitope). The IC_{50} values of standard peptides used to determine binding affinities for Class I peptides are shown in Table IV. The IC_{50} values of standard peptides used to determine binding affinities for Class II peptides are shown in Table V. The peptides used as standards for the binding assays described herein are examples of standards; alternative standard peptides can also be used when performing such an analysis.

To obtain the peptide epitope sequences listed in each Table, protein sequence data from fourteen HCV isolates were evaluated for the presence of the designated supermotif or motif. The fourteen strains include HPCCGAA, HPCPLYPRE, HCV-H-CMR, HCV-J1, HPCGENANTI, HPCGENOM, HPCHUMR, HPCJCG, HPCJTA, HCV-J483, HCV-JK1, HCV-N, HPCPOLP, and HCV-J8. Peptide epitopes were additionally evaluated on the basis of their conservancy among these fourteen strains. A criterion for conservancy requires that the entire sequence of an HLA class I binding peptide be totally conserved in 79% of the sequences available for a specific protein. Similarly, a criterion for conservancy requires that the entire 9-mer core region of an HLA class II binding

peptide be totally conserved in 79% of the sequences available for a specific protein. The percent conservancy of the selected peptide epitopes is indicated on the Tables. The frequency, *i.e.* the number of strains of the fourteen strains in which the totally conserved peptide sequence was identified, is also shown. The "position" column in the Tables designates the amino acid position of the HCV polyprotein that corresponds to the first amino acid residue of the epitope. The "number of amino acids" indicates the number of residues in the epitope sequence.

HLA Class I Motifs Indicative of CTL Inducing Peptide Epitopes:

The primary anchor residues of the HLA class I peptide epitope supermotifs and motifs delineated below are summarized in Table I. The HLA class I motifs set out in Table I(a) are those most particularly relevant to the invention claimed here. Primary and secondary anchor positions are summarized in Table II. Allele-specific HLA molecules that comprise HLA class I supertype families are listed in Table VI.

IV.D.1. HLA-A1 supermotif

The HLA-A1 supermotif is characterized by the presence in peptide ligands of a small (T or S) or hydrophobic (L, I, V, or M) primary anchor residue in position 2, and an aromatic (Y, F, or W) primary anchor residue at the C-terminal position of the epitope. The corresponding family of HLA molecules that bind to the A1 supermotif (*i.e.*, the HLA-A1 supertype) includes at least A*0101, A*2601, A*2602, A*2501, and A*3201 (*see*, *e.g.*, DiBrino, M. *et al.*, *J. Immunol.* 151:5930, 1993; DiBrino, M. *et al.*, *J. Immunol.* 152:620, 1994; Kondo, A. *et al.*, *Immunogenetics* 45:249, 1997). Other allele-specific HLA molecules predicted to be members of the A1 superfamily are shown in Table VI. Peptides binding to each of the individual HLA proteins can be modulated by substitutions at primary and/or secondary anchor positions, preferably choosing respective residues specified for the supermotif.

Peptide epitopes that comprise the A1 supermotif are set forth in Table VII.

IV.D.2. HLA-A2 supermotif

Primary anchor specificities for allele-specific HLA-A2.1 molecules (Falk *et al.*, *Nature* 351:290-296, 1991; Hunt *et al.*, *Science* 255:1261-1263, 1992; Parker *et al.*, *J. Immunol.* 149:3580-3587, 1992) and cross-reactive binding within the HLA A2 family (Fruci *et al.*, *Human Immunol.* 38:187-192, 1993; Tanigaki *et al.*, *Human Immunol.*

39:155-162, 1994) have been described. The present inventors have defined additional primary anchor residues that determine cross-reactive binding to multiple allele-specific HLA A2 molecules (Ruppert *et al.*, *Cell* 74:929-937, 1993; Del Guercio *et al.*, *J. Immunol.* 154:685-693, 1995; Kast *et al.*, *J. Immunol.* 152:3904-3912, 1994). The HLA-A2 supermotif comprises peptide ligands with L, I, V, M, A, T, or Q as a primary anchor residue at position 2 and L, I, V, M, A, or T as a primary anchor residue at the C-terminal position of the epitope.

The corresponding family of HLA molecules (*i.e.*, the HLA-A2 supertype that binds these peptides) is comprised of at least: A*0201, A*0202, A*0203, A*0204, A*0205, A*0206, A*0207, A*0209, A*0214, A*6802, and A*6901. Other allele-specific HLA molecules predicted to be members of the A2 superfamily are shown in Table VI. As explained in detail below, binding to each of the individual allele-specific HLA molecules can be modulated by substitutions at the primary anchor and/or secondary anchor positions, preferably choosing respective residues specified for the supermotif.

Peptide epitopes that comprise an A2 supermotif are set forth inTable VIII. The motifs comprising the primary anchor residues V, A, T, or Q at position 2 and L, I, V, A, or T at the C-terminal position are those most particularly relevant to the invention claimed herein.

IV.D.3. HLA-A3 supermotif

The HLA-A3 supermotif is characterized by the presence in peptide ligands of A, L, I, V, M, S, or, T as a primary anchor at position 2, and a positively charged residue, R or K, at the C-terminal position of the epitope (*e.g.*, in position 9 of 9-mers). Exemplary members of the corresponding family of HLA molecules (the HLA-A3 supertype) that bind the A3 supermotif include at least A*0301, A*1101, A*3101, A*3301, and A*6801. Other allele-specific HLA molecules predicted to be members of the A3 superfamily are shown in Table VI. As explained in detail below, peptide binding to each of the individual allele-specific HLA proteins can be modulated by substitutions of amino acids at the primary and/or secondary anchor positions of the peptide, preferably choosing respective residues specified for the supermotif.

Peptide epitopes that comprise the A3 supermotif are set forth in Table IX.

IV.D.4. HLA-A24 supermotif

The HLA-A24 supermotif is characterized by the presence in peptide ligands of an aromatic (F, W, or Y) or hydrophobic aliphatic (L, I, V, M, or T) residue as a primary anchor in position 2, and Y, F, W, L, I, or M as primary anchor at the C-terminal position of the epitope. The corresponding family of HLA molecules that bind to the A24 supermotif (*i.e.*, the A24 supertype) includes at least A*2402, A*3001, and A*2301. Other allele-specific HLA molecules predicted to be members of the A24 superfamily are shown in Table VI. Peptide binding to each of the allele-specific HLA molecules can be modulated by substitutions at primary and/or secondary anchor positions, preferably choosing respective residues specified for the supermotif.

Peptide epitopes that comprise the A24 supermotif are set forth in Table X.

IV.D.5. HLA-B7 supermotif

The HLA-B7 supermotif is characterized by peptides bearing proline in position 2 as a primary anchor, and a hydrophobic or aliphatic amino acid (L, I, V, M, A, F, W, or Y) as the primary anchor at the C-terminal position of the epitope. The corresponding family of HLA molecules that bind the B7 supermotif (*i.e.*, the HLA-B7 supertype) is comprised of at least twenty six HLA-B proteins including: B*0702, B*0703, B*0704, B*0705, B*1508, B*3501, B*3502, B*3503, B*3504, B*3505, B*3506, B*3507, B*3508, B*5101, B*5102, B*5103, B*5104, B*5105, B*5301, B*5401, B*5501, B*5502, B*5601, B*5602, B*6701, and B*7801 (*see, e.g.*, Sidney, *et al.*, *J. Immunol.* 154:247, 1995; Barber, *et al.*, *Curr. Biol.* 5:179, 1995; Hill, *et al.*, *Nature* 360:434, 1992; Rammensee, *et al.*, *Immunogenetics* 41:178, 1995). Other allele-specific HLA molecules predicted to be members of the B7 superfamily are shown in Table VI. As explained in detail below, peptide binding to each of the individual allele-specific HLA proteins can be modulated by substitutions at the primary and/or secondary anchor positions of the peptide, preferably choosing respective residues specified for the supermotif.

Peptide epitopes that comprise the B7 supermotif are set forth in Table XI.

IV.D.6. HLA-B27 supermotif

The HLA-B27 supermotif is characterized by the presence in peptide ligands of a positively charged (R, H, or K) residue as a primary anchor at position 2, and a hydrophobic (F, Y, L, W, M, I, A, or V) residue as a primary anchor at the C-terminal position of the epitope. Exemplary members of the corresponding family of HLA

molecules that bind to the B27 supermotif (*i.e.*, the B27 supertype) include at least B*1401, B*1402, B*1509, B*2702, B*2703, B*2704, B*2705, B*2706, B*3801, B*3901, B*3902, and B*7301. Other allele-specific HLA molecules predicted to be members of the B27 superfamily are shown in Table VI. Peptide binding to each of the allele-specific HLA molecules can be modulated by substitutions at primary and/or secondary anchor positions, preferably choosing respective residues specified for the supermotif.

Peptide epitopes that comprise the B27 supermotif are set forth in Table XII.

IV.D.7. HLA-B44 supermotif

The HLA-B44 supermotif is characterized by the presence in peptide ligands of negatively charged (D or E) residues as a primary anchor in position 2, and hydrophobic residues (F, W, Y, L, I, M, V, or A) as a primary anchor at the C-terminal position of the epitope. Exemplary members of the corresponding family of HLA molecules that bind to the B44 supermotif (*i.e.*, the B44 supertype) include at least: B*1801, B*1802, B*3701, B*4001, B*4002, B*4006, B*4402, B*4403, and B*4006. Peptide binding to each of the allele-specific HLA molecules can be modulated by substitutions at primary and/or secondary anchor positions; preferably choosing respective residues specified for the supermotif.

IV.D.8. HLA-B58 supermotif

The HLA-B58 supermotif is characterized by the presence in peptide ligands of a small aliphatic residue (A, S, or T) as a primary anchor residue at position 2, and an aromatic or hydrophobic residue (F, W, Y, L, I, V, M, or A) as a primary anchor residue at the C-terminal position of the epitope. Exemplary members of the corresponding family of HLA molecules that bind to the B58 supermotif (*i.e.*, the B58 supertype) include at least: B*1516, B*1517, B*5701, B*5702, and B*5801. Other allele-specific HLA molecules predicted to be members of the B58 superfamily are shown in Table VI. Peptide binding to each of the allele-specific HLA molecules can be modulated by substitutions at primary and/or secondary anchor positions, preferably choosing respective residues specified for the supermotif.

Peptide epitopes that comprise the B58 supermotif are set forth in Table XIII.

IV.D.9. HLA-B62 supermotif

The HLA-B62 supermotif is characterized by the presence in peptide ligands of the polar aliphatic residue Q or a hydrophobic aliphatic residue (L, V, M, I, or P) as a primary anchor in position 2, and a hydrophobic residue (F, W, Y, M, I, V, L, or A) as a primary anchor at the C-terminal position of the epitope. Exemplary members of the corresponding family of HLA molecules that bind to the B62 supermotif (*i.e.*, the B62 supertype) include at least: B*1501, B*1502, B*1513, and B5201. Other allele-specific HLA molecules predicted to be members of the B62 superfamily are shown in Table VI. Peptide binding to each of the allele-specific HLA molecules can be modulated by substitutions at primary and/or secondary anchor positions, preferably choosing respective residues specified for the supermotif.

Peptide epitopes that comprise the B62 supermotif are set forth in Table XIV.

IV.D.10. HLA-A1 motif

The HLA-A1 motif is characterized by the presence in peptide ligands of T, S, or M as a primary anchor residue at position 2 and the presence of Y as a primary anchor residue at the C-terminal position of the epitope. An alternative allele-specific A1 motif is characterized by a primary anchor residue at position 3 rather than position 2. This motif is characterized by the presence of D, E, A, or S as a primary anchor residue in position 3, and a Y as a primary anchor residue at the C-terminal position of the epitope. Peptide binding to HLA A1 can be modulated by substitutions at primary and/or secondary anchor positions, preferably choosing respective residues specified for the motif.

Peptide epitopes that comprise either A1 motif are set forth in Table XV. The epitopes comprising T, S, or M at position 2 and Y at the C-terminal position are also included in the listing of HLA-A1 supermotif-bearing peptide epitopes listed in Table VII.

IV.D.11. HLA-A*0201 motif

An HLA-A2*0201 motif was first determined to be characterized by the presence in peptide ligands of L or M as a primary anchor residue in position 2, and L or V as a primary anchor residue at the C-terminal position of a 9-residue peptide (Falk *et al.*, *Nature* 351:290-296, 1991). The A*0201 motif was also determined to further comprise an I at position 2 and I or A at the C-terminal position of a nine amino acid peptide (Hunt

et al., Science 255:1261-1263, March 6, 1992; Parker *et al.*, *J. Immunol.* 149:3580-3587, 1992). Subsequently, the A*0201 allele-specific motif has been defined by the present inventors to additionally comprise V, A, T, or Q as a primary anchor residue at position 2, and M as a primary anchor residue at the C-terminal position of the epitope. Additionally, the A*0201 allele-specific motif has been found to comprise a T at the C-terminal position (Kast *et al.*, *J. Immunol.* 152:3904-3912, 1994). Thus, the HLA-A*0201 motif comprises peptide ligands with L, I, V, M, A, T, or Q as primary anchor residues at position 2 and L, I, V, M, A, or T as a primary anchor residue at the C-terminal position of the epitope. The preferred and tolerated residues that characterize the primary anchor positions of the HLA-A*0201 motif are identical to the residues describing the A2 supermotif. (For reviews of relevant data, *see*, *e.g.*, Del Guercio *et al.*, *J. Immunol.* 154:685-693, 1995; Ruppert *et al.*, *Cell* 74:929-937, 1993; Sidney *et al.*, *Immunol. Today* 17:261-266, 1996; Sette and Sidney, *Curr. Opin. in Immunol.* 10:478-482, 1998). Secondary anchor residues that characterize the A*0201 motif have additionally been defined as disclosed herein. These are disclosed in Table II. Peptide binding to HLA-A*0201 molecules can be modulated by substitutions at primary and/or secondary anchor positions, preferably choosing respective residues specified for the motif.

Peptide epitopes that comprise an A*0201 motif are set forth in Table VIII. The A*0201 motifs comprising the primary anchor residues V, A, T, or Q at position 2 and L, I, V, A, or T at the C-terminal position are those most particularly relevant to the invention claimed herein.

IV.D.12. HLA-A3 motif

The HLA-A3 motif is characterized by the presence in peptide ligands of L, M, V, I, S, A, T, F, C, G, or D as a primary anchor residue at position 2, and the presence of K, Y, R, H, F, or A as a primary anchor residue at the C-terminal position of the epitope. Peptide binding to HLA-A3 can be modulated by substitutions at primary and/or secondary anchor positions, preferably choosing respective residues specified for the motif.

The A3 supermotif primary anchor residues comprise a subset of the A3- and A11-allele specific motif primary anchor residues. Peptide epitopes that comprise the A3 motif are set forth inTable XVI. Those peptide epitopes that also comprise the A3 supermotif are also listed in Table IX.

IV.D.13. HLA-A11 m tif

The HLA-A11 motif is characterized by the presence in peptide ligands of V, T, M, L, I, S, A, G, N, C, D, or F as a primary anchor residue in position 2, and K, R, Y, or H as a primary anchor residue at the C-terminal position of the epitope. Peptide binding to HLA-A11 can be modulated by substitutions at primary and/or secondary anchor positions, preferably choosing respective residues specified for the motif.

Peptide epitopes that comprise the A11 motif are set forth in Table XVII; peptide epitopes comprising the A3 allele-specific motif are also present in this Table because of the overlap between the A3 and A11 motif primary anchor specificities. Further, those peptide epitopes that comprise the A3 supermotif are also listed in Table IX.

IV.D.14. HLA-A24 motif

The HLA-A24 motif is characterized by the presence in peptide ligands of Y, F, W, or M as a primary anchor residue in position 2, and F, L, I, or W as a primary anchor residue at the C-terminal position of the epitope. Peptide binding to HLA-A24 molecules can be modulated by substitutions at primary and/or secondary anchor positions; preferably choosing respective residues specified for the motif.

Peptide epitopes that comprise the A24 motif are set forth inTable XVIII. These epitopes are also listed in Table X, which sets forth HLA-A24-supermotif-bearing peptide epitopes, as the primary anchor residues characterizing the A24 allele-specific motif comprise a subset of the A24 supermotif primary anchor residues.

HLA Class II Binding Motifs

The primary and secondary anchor residues of the HLA class II peptide epitope supermotifs and motifs delineated below are summarized in Table III.

IV.D.15. HLA DR-1-4-7 supermotif

Motifs have also been identified for peptides that bind to three common HLA class II allele-specific HLA molecules: HLA DRB1*0401, DRB1*0101, and DRB1*0701. Collectively, the common residues from these motifs delineate the HLA DR-1-4-7 supermotif. Peptides that bind to these DR molecules carry a supermotif characterized by a large aromatic or hydrophobic residue (Y, F, W, L, I, V, or M) as a primary anchor residue in position 1, and a small, non-charged residue (S, T, C, A, P, V,

I, L, or M) as a primary anchor residue in position 6 of a 9-mer core region. Allele-specific secondary effects and secondary anchors for each of these HLA types have also been identified. These are set forth in Table III. Peptide binding to HLA- DRB1*0401, DRB1*0101, and/or DRB1*0701 can be modulated by substitutions at primary and/or secondary anchor positions, preferably choosing respective residues specified for the supermotif.

Conserved peptide epitopes *i.e.*, conserved in ≥79% (≥11/14) of the HCV strains used for the present analysis, may be described as corresponding to epitopes containing a nine residue core comprising the DR-1-4-7 supermotif, and in which the 9 residue core is conserved in ≥79% (wherein position 1 of the motif is at position 1 of the nine residue core). Conserved 9-mer core regions are set forth in Table XIXa. Respective exemplary peptide epitopes of 15 amino acid residues in length, each of which comprise a conserved nine residue core, are also shown in section "a" of the Table. Cross-reactive binding data for exemplary 15-residue supermotif-bearing peptides are shown in Table XIXb.

IV.D.16. HLA DR3 motifs

Two alternative motifs (*i.e.*, submotifs) characterize peptide epitopes that bind to HLA-DR3 molecules. In the first motif (submotif DR3A) a large, hydrophobic residue (L, I, V, M, F, or Y) is present in anchor position 1 of a 9-mer core, and D is present as an anchor at position 4, towards the carboxyl terminus of the epitope. As in other class II motifs, core position 1 may or may not occupy the peptide N-terminal position.

The alternative DR3 submotif provides for lack of the large, hydrophobic residue at anchor position 1, and/or lack of the negatively charged or amide-like anchor residue at position 4, by the presence of a positive charge at position 6 towards the carboxyl terminus of the epitope. Thus, for the alternative allele-specific DR3 motif (submotif DR3B): L, I, V, M, F, Y, A, or Y is present at anchor position 1; D, N, Q, E, S, or T is present at anchor position 4; and K, R, or H is present at anchor position 6. Peptide binding to HLA-DR3 can be modulated by substitutions at primary and/or secondary anchor positions, preferably choosing respective residues specified for the motif.

Conserved 9-mer core regions (*i.e.*, those sequences that are conserved in at least 79% of the 14 HCV strains used for the analysis) corresponding to a nine residue sequence comprising the DR3A submotif (wherein position 1 of the motif is at position 1 of the nine residue core) are set forth in Table XXa. Respective exemplary peptide

epitopes of 15 amino acid residues in length, each of which comprise a conserved nine residue core, are also shown in Table XXa. Table XXb shows binding data of exemplary DR3 submotif A-bearing peptides.

Conserved 9-mer core regions (*i.e.*, those that are at least 79% conserved in the 14 HCV strains used for the analysis) comprising the DR3B submotif and respective exemplary 15-mer peptides comprising the DR3 submotif-B epitope are set forth in Table XXc. Table XXd shows binding data of exemplary DR3 submotif B-bearing peptides.

Each of the HLA class I or class II peptide epitopes set out in the Tables herein are deemed singly to be an inventive aspect of this application. Further, it is also an inventive aspect of this application that each peptide epitope may be used in combination with any other peptide epitope.

IV.E. Enhancing Population Coverage of the Vaccine

Vaccines that have broad population coverage are preferred because they are more commercially viable and generally applicable to the most people. Broad population coverage can be obtained using the peptides of the invention (and nucleic acid compositions that encode such peptides) through selecting peptide epitopes that bind to HLA alleles which, when considered in total, are present in most of the population. Table XXI lists the overall frequencies of the HLA class I supertypes in various ethnicities (Table XXIa) and the combined population coverage achieved by the A2-, A3-, and B7-supertypes (Table XXIb). The A2-, A3-, and B7 supertypes are each present on the average of over 40% in each of these five major ethnic groups. Coverage in excess of 80% is achieved with a combination of these supermotifs. These results suggest that effective and non-ethnically biased population coverage is achieved upon use of a limited number of cross-reactive peptides. Although the population coverage reached with these three main peptide specificities is high, coverage can be expanded to reach 95% population coverage and above, and more easily achieve truly multispecific responses upon use of additional supermotif or allele-specific motif bearing peptides.

The B44-, A1-, and A24-supertypes are present, on average, in a range from 25% to 40% in these major ethnic populations (Table XXIa). While less prevalent overall, the B27-, B58-, and B62 supertypes are each present with a frequency >25% in at least one major ethnic group (Table XXIa). Table XXIb summarizes the estimated prevalence of combinations of HLA supertypes that have been identified in five major ethnic groups.

The incremental coverage obtained by the inclusion of A1,- A24-, and B44-supertypes to the A2, A3, and B7 coverage, or all of the supertypes described herein, is shown.

The data presented herein, together with the previous definition of the A2-, A3-, and B7-supertypes, indicates that all antigens, with the possible exception of A29, B8, and B46, can be classified into a total of nine HLA supertypes. By including epitopes from the six most frequent supertypes, an average population coverage of 99% is obtained for five major ethnic groups..

IV.F. Immune Response-Stimulating Peptide Analogs

In general, CTL and HTL responses are not directed against all possible epitopes. Rather, they are restricted to a few "immunodominant" determinants (Zinkernagel, *et al., Adv. Immunol.* 27:5159, 1979; Bennink, *et al., J. Exp. Med.* 168:19351939, 1988; Rawle, *et al., J. Immunol.* 146:3977-3984, 1991). It has been recognized that immunodominance (Benacerraf, *et al., Science* 175:273-279, 1972) could be explained by either the ability of a given epitope to selectively bind a particular HLA protein (determinant selection theory) (Vitiello, *et al., J. Immunol.* 131:1635, 1983); Rosenthal, *et al., Nature* 267:156-158, 1977), or to be selectively recognized by the existing TCR (T cell receptor) specificities (repertoire theory) (Klein, J., IMMUNOLOGY, THE SCIENCE OF SELFNONSELF DISCRIMINATION, John Wiley & Sons, New York, pp. 270-310, 1982). It has been demonstrated that additional factors, mostly linked to processing events, can also play a key role in dictating, beyond strict immunogenicity, which of the many potential determinants will be presented as immunodominant (Sercarz, *et al., Annu. Rev. Immunol.* 11:729-766, 1993).

The concept of dominance and subdominance is relevant to immunotherapy of both infectious diseases and cancer. For example, in the course of chronic viral disease, recruitment of subdominant epitopes can be important for successful clearance of the infection, especially if dominant CTL or HTL specificities have been inactivated by functional tolerance, suppression, mutation of viruses and other mechanisms (Franco, *et al., Curr. Opin. Immunol.* 7:524-531, 1995). In the case of cancer and tumor antigens, CTLs recognizing at least some of the highest binding affinity peptides might be functionally inactivated. Lower binding affinity peptides are preferentially recognized at these times, and may therefore be preferred in therapeutic or prophylactic anti-cancer vaccines.

In particular, it has been noted that a significant number of epitopes derived from known non-viral tumor associated antigens (TAA) bind HLA class I with intermediate affinity (IC_{50} in the 50-500 nM range). For example, it has been found that 8 of 15 known TAA peptides recognized by tumor infiltrating lymphocytes (TIL) or CTL bound in the 50-500 nM range. (These data are in contrast with estimates that 90% of known viral antigens were bound by HLA class I molecules with IC_{50} of 50 nM or less, while only approximately 10% bound in the 50-500 nM range (Sette, *et al., J. Immunol.*, 153:558-5592, 1994). In the cancer setting this phenomenon is probably due to elimination or functional inhibition of the CTL recognizing several of the highest binding peptides, presumably because of T cell tolerization events.

Without intending to be bound by theory, it is believed that because T cells to dominant epitopes may have been clonally deleted, selecting subdominant epitopes may allow existing T cells to be recruited, which will then lead to a therapeutic or prophylactic response. However, the binding of HLA molecules to subdominant epitopes is often less vigorous than to dominant ones. Accordingly, there is a need to be able to modulate the binding affinity of particular immunogenic epitopes for one or more HLA molecules, and thereby to modulate the immune response elicited by the peptide, for example to prepare analog peptides which elicit a more vigorous response. This ability would greatly enhance the usefulness of peptide-based vaccines and therapeutic agents.

Although peptides with suitable cross-reactivity among all alleles of a superfamily are identified by the screening procedures described above, cross-reactivity is not always as complete as possible, and in certain cases procedures to increase cross-reactivity of peptides can be useful; moreover, such procedures can also be used to modify other properties of the peptides such as binding affinity or peptide stability. Having established the general rules that govern cross-reactivity of peptides for HLA alleles within a given motif or supermotif, modification (*i.e.*, analoging) of the structure of peptides of particular interest in order to achieve broader (or otherwise modified) HLA binding capacity can be performed. More specifically, peptides which exhibit the broadest cross-reactivity patterns, can be produced in accordance with the teachings herein. The present concepts related to analog generation are set forth in greater detail in co-pending U.S.S.N. 09/226,775 filed 1/6/99.

In brief, the strategy employed utilizes the motifs or supermotifs which correlate with binding to certain HLA molecules. The motifs or supermotifs are defined by having primary anchors, and in many cases secondary anchors. Analog peptides can be created

by substituting amino acid residues at primary anchor, secondary anchor, or at primary and secondary anchor positions. Generally, analogs are made for peptides that already bear a motif or supermotif. Preferred secondary anchor residues of supermotifs and motifs that have been defined for HLA class I and class II binding peptides are shown in Tables II and III, respectively.

For a number of the motifs or supermotifs in accordance with the invention, residues are defined which are deleterious to binding to allele-specific HLA molecules or members of HLA supertypes that bind the respective motif or supermotif (Tables II and III). Accordingly, removal of such residues that are detrimental to binding can be performed in accordance with the present invention. For example, in the case of the A3 supertype, when all peptides that have such deleterious residues are removed from the population of analyzed peptides, the incidence of cross-reactivity increases from 22% to 37% (*see, e.g.*, Sidney, J. *et al., Hu. Immunol.* 45:79, 1996). Thus, one strategy to improve the cross-reactivity of peptides within a given supermotif is simply to delete one or more of the deleterious residues present within a peptide and substitute a small "neutral" residue such as Ala (that may not influence T cell recognition of the peptide). An enhanced likelihood of cross-reactivity is expected if, together with elimination of detrimental residues within a peptide, "preferred" residues associated with high affinity binding to an allele-specific HLA molecule or to multiple HLA molecules within a superfamily are inserted.

To ensure that an analog peptide, when used as a vaccine, actually elicits a CTL response to the native epitope *in vivo* (or, in the case of class II epitopes, elicits helper T cells that cross-react with the wild type peptides), the analog peptide may be used to immunize T cells *in vitro* from individuals of the appropriate HLA allele. Thereafter, the immunized cells' capacity to induce lysis of wild type peptide sensitized target cells is evaluated. It will be desirable to use as antigen presenting cells, cells that have been either infected, or transfected with the appropriate genes, or, in the case of class II epitopes only, cells that have been pulsed with whole protein antigens, to establish whether endogenously produced antigen is also recognized by the relevant T cells.

Another embodiment of the invention is to create analogs of weak binding peptides, to thereby ensure adequate numbers of cross-reactive cellular binders. Class I binding peptides exhibiting binding affinities of 500-5000 nM, and carrying an acceptable but suboptimal primary anchor residue at one or both positions can be "fixed" by

substituting preferred anchor residues in accordance with the respective supertype. The analog peptides can then be tested for crossbinding activity.

Another embodiment for generating effective peptide analogs involves the substitution of residues that have an adverse impact on peptide stability or solubility in, *e.g.*, a liquid environment. This substitution may occur at any position of the peptide epitope. For example, a cysteine (C) can be substituted out in favor of α-amino butyric acid. Due to its chemical nature, cysteine has the propensity to form disulfide bridges and sufficiently alter the peptide structurally so as to reduce binding capacity. Substituting α-amino butyric acid for C not only alleviates this problem, but actually improves binding and crossbinding capability in certain instances (*see*, *e.g.*, the review by Sette *et al.*, In: Persistent Viral Infections, Eds. R. Ahmed and I. Chen, John Wiley & Sons, England, 1999). Substitution of cysteine with α-amino butyric acid may occur at any residue of a peptide epitope, *i.e.* at either anchor or non-anchor positions.

Representative analog peptides are set forth in Table XXII. The Table indicates the length and sequence of the analog peptide as well as the motif or supermotif, if appropriate. The information in the "Fixed Nomenclature" column indicates the residues substituted at the indicated position numbers for the respective analog.

IV.G. Computer Screening of Protein Sequences from Disease-Related Antigens for Supermotif- or Motif-Bearing Peptides

In order to identify supermotif- or motif-bearing epitopes in a target antigen, a native protein sequence, *e.g.*, a tumor-associated antigen, or sequences from an infectious organism, or a donor tissue for transplantation, is screened using a means for computing, such as an intellectual calculation or a computer, to determine the presence of a supermotif or motif within the sequence. The information obtained from the analysis of native peptide can be used directly to evaluate the status of the native peptide or may be utilized subsequently to generate the peptide epitope.

Computer programs that allow the rapid screening of protein sequences for the occurrence of the subject supermotifs or motifs are encompassed by the present invention; as are programs that permit the generation of analog peptides. These programs are implemented to analyze any identified amino acid sequence or operate on an unknown sequence and simultaneously determine the sequence and identify motif-bearing epitopes thereof; analogs can be simultaneously determined as well. Generally, the identified sequences will be from a pathogenic organism or a tumor-associated peptide. For

example, the target molecules considered herein include, without limitation, the core, S, E1, NS1/E2, NS2, NS3, NS4, and NS5 regions of HCV.

In cases where the sequence of multiple variants of the same target protein are available, peptides may also be selected on the basis of their conservancy. A presently preferred criterion for conservancy defines that the entire sequence of an HLA class I binding peptide or the entire 9-mer core of a class II binding peptide, be totally (*i.e.*, 100%) conserved in at least 79% of the sequences evaluated for a specific protein. This definition of conservancy has been employed herein; although, as appreciated by those in the art, lower or higher degrees of conservancy can be employed as appropriate for a given antigenic target.

It is important that the selection criteria utilized for prediction of peptide binding are as accurate as possible, to correlate most efficiently with actual binding. Prediction of peptides that bind, for example, to HLA-A*0201, on the basis of the presence of the appropriate primary anchors, is positive at about a 30% rate (see, *e.g.*, Ruppert, J. *et al. Cell* 74:929, 1993). However, by extensively analyzing peptide-HLA binding data disclosed herein, data in related patent applications, and data in the art, the present inventors have developed a number of allele-specific polynomial algorithms that dramatically increase the predictive value over identification on the basis of the presence of primary anchor residues alone. These algorithms take into account not only the presence or absence of primary anchors, but also consider the positive or deleterious presence of secondary anchor residues (to account for the impact of different amino acids at different positions). The algorithms are essentially based on the premise that the overall affinity (or ΔG) of peptide-HLA interactions can be approximated as a linear polynomial function of the type:

$$\Delta G = a_{1i} \times a_{2i} \times a_{3i} \ldots \times a_{ni}$$

where a_{ji} is a coefficient that represents the effect of the presence of a given amino acid (j) at a given position (i) along the sequence of a peptide of n amino acids. An important assumption of this method is that the effects at each position are essentially independent of each other. This assumption is justified by studies that demonstrated that peptides are bound to HLA molecules and recognized by T cells in essentially an extended conformation. Derivation of specific algorithm coefficients has been described, for example, in Gulukota, K. *et al.*, *J. Mol. Biol.* 267:1258, 1997.

Additional methods to identify preferred peptide sequences, which also make use of specific motifs, include the use of neural networks and molecular modeling programs (*see, e.g.*, Milik *et al.*, *Nature Biotechnology* 16:753, 1998; Altuvia *et al.*, *Hum. Immunol.* 58:1, 1997; Altuvia *et al*, *J. Mol. Biol.* 249:244, 1995; Buus, S. *Curr. Opin. Immunol.* 11:209-213, 1999; Brusic, V. *et al.*, *Bioinformatics* 14:121-130, 1998; Parker *et al.*, *J. Immunol.* 152:163, 1993; Meister *et al.*, *Vaccine* 13:581, 1995; Hammer *et al.*, *J. Exp. Med.* 180:2353, 1994; Sturniolo *et al.*, *Nature Biotechnol.* 17:555 1999).

For example, it has been shown that in sets of A*0201 motif-bearing peptides containing at least one preferred secondary anchor residue while avoiding the presence of any deleterious secondary anchor residues, 69% of the peptides will bind A*0201 with an IC_{50} less than 500 nM (Ruppert, J. *et al. Cell* 74:929, 1993). These algorithms are also flexible in that cut-off scores may be adjusted to select sets of peptides with greater or lower predicted binding properties, as desired.

In utilizing computer screening to identify peptide epitopes, a protein sequence or translated sequence may be analyzed using software developed to search for motifs, for example the "FINDPATTERNS' program (Devereux, *et a*l. *Nucl. Acids Res.* 12:387-395, 1984) or MotifSearch 1.4 software program (D. Brown, San Diego, CA) to identify potential peptide sequences containing appropriate HLA binding motifs. The identified peptides can be scored using customized polynomial algorithms to predict their capacity to bind specific HLA class I or class II alleles. As appreciated by one of ordinary skill in the art, a large array of computer programming software and hardware options are available in the relevant art which can be employed to implement the motifs of the invention in order to evaluate (*e.g.*, without limitation, to identify epitopes, identify epitope concentration per peptide length, or to generate analogs) known or unknown peptide sequences.

In accordance with the procedures described above, HCV peptide epitopes and analogs thereof that are able to bind HLA supertype groups or allele-specific HLA molecules have been identified (Tables VII-XX; Table XXII).

IV.H. Preparation of Peptide Epitopes

Peptides in accordance with the invention can be prepared synthetically, by recombinant DNA technology or chemical synthesis, or from natural sources such as native tumors or pathogenic organisms. Peptide epitopes may be synthesized individually or as polyepitopic peptides. Although the peptide will preferably be substantially free of

other naturally occurring host cell proteins and fragments thereof, in some embodiments the peptides may be synthetically conjugated to native fragments or particles.

The peptides in accordance with the invention can be a variety of lengths, and either in their neutral (uncharged) forms or in forms which are salts. The peptides in accordance with the invention are either free of modifications such as glycosylation, side chain oxidation, or phosphorylation; or they contain these modifications, subject to the condition that modifications do not destroy the biological activity of the peptides as described herein.

The peptides of the invention can be prepared in a wide variety of ways. For the preferred relatively short size, the peptides can be synthesized in solution or on a solid support in accordance with conventional techniques. Various automatic synthesizers are commercially available and can be used in accordance with known protocols. (*See*, for example, Stewart & Young, SOLID PHASE PEPTIDE SYNTHESIS, 2D. ED., Pierce Chemical Co., 1984). Further, individual peptide epitopes can be joined using chemical ligation to produce larger peptides that are still within the bounds of the invention.

Alternatively, recombinant DNA technology can be employed wherein a nucleotide sequence which encodes an immunogenic peptide of interest is inserted into an expression vector, transformed or transfected into an appropriate host cell and cultivated under conditions suitable for expression. These procedures are generally known in the art, as described generally in Sambrook *et al.,* MOLECULAR CLONING, A LABORATORY MANUAL, Cold Spring Harbor Press, Cold Spring Harbor, New York (1989). Thus, recombinant polypeptides which comprise one or more peptide sequences of the invention can be used to present the appropriate T cell epitope.

The nucleotide coding sequence for peptide epitopes of the preferred lengths contemplated herein can be synthesized by chemical techniques, for example, the phosphotriester method of Matteucci, *et al., J. Am. Chem. Soc.* 103:3185 (1981). Peptide analogs can be made simply by substituting the appropriate and desired nucleic acid base(s) for those that encode the native peptide sequence; exemplary nucleic acid substitutions are those that encode an amino acid defined by the motifs/supermotifs herein. The coding sequence can then be provided with appropriate linkers and ligated into expression vectors commonly available in the art, and the vectors used to transform suitable hosts to produce the desired fusion protein. A number of such vectors and suitable host systems are now available. For expression of the fusion proteins, the coding sequence will be provided with operably linked start and stop codons, promoter and

terminator regions and usually a replication system to provide an expression vector for expression in the desired cellular host. For example, promoter sequences compatible with bacterial hosts are provided in plasmids containing convenient restriction sites for insertion of the desired coding sequence. The resulting expression vectors are transformed into suitable bacterial hosts. Of course, yeast, insect or mammalian cell hosts may also be used, employing suitable vectors and control sequences.

It is often preferable that the peptide epitope be as small as possible while still maintaining substantially all of the immunologic activity of the native protein. When possible, it may be desirable to optimize HLA class I binding peptide epitopes of the invention to a length of about 8 to about 13 amino acid residues, preferably 9 to 10. HLA class II binding peptide epitopes may be optimized to a length of about 6 to about 30 amino acids in length, preferably to between about 13 and about 20 residues. Preferably, the peptide epitopes are commensurate in size with endogenously processed pathogen-derived peptides or tumor cell peptides that are bound to the relevant HLA molecules, however, the identification and preparation of peptides of other lengths can also be carried out using the techniques described herein.

In alternative embodiments, peptides of the invention can be linked as a polyepitopic peptide, or as a minigene that encodes a polyepitopic peptide.

In another embodiment, it is preferred to identify native peptide regions that contain a high concentration of class I and/or class II epitopes. Such a sequence is generally selected on the basis that it contains the greatest number of epitopes per amino acid length. It is to be appreciated that epitopes can be present in a frame-shifted manner, *e.g.* a 10 amino acid long peptide could contain two 9 amino acid long epitopes and one 10 amino acid long epitope; upon intracellular processing, each epitope can be exposed and bound by an HLA molecule upon administration of such a peptide. This larger, preferably multi-epitopic, peptide can be generated synthetically, recombinantly, or via cleavage from the native source.

IV.I. Assays to Detect T-Cell Responses

Once HLA binding peptides are identified, they can be tested for the ability to elicit a T-cell response. The preparation and evaluation of motif-bearing peptides are described in PCT publications WO 94/20127 and WO 94/03205. Briefly, peptides comprising epitopes from a particular antigen are synthesized and tested for their ability to bind to the appropriate HLA proteins. These assays may involve evaluating the

binding of a peptide of the invention to purified HLA class I molecules in relation to the binding of a radioiodinated reference peptide. Alternatively, cells expressing empty class I molecules (*i.e.* lacking peptide therein) may be evaluated for peptide binding by immunofluorescent staining and flow microfluorimetry. Other assays that may be used to evaluate peptide binding include peptide-dependent class I assembly assays and/or the inhibition of CTL recognition by peptide competition. Those peptides that bind to the class I molecule, typically with an affinity of 500 nM or less, are further evaluated for their ability to serve as targets for CTLs derived from infected or immunized individuals, as well as for their capacity to induce primary *in vitro* or *in vivo* CTL responses that can give rise to CTL populations capable of reacting with selected target cells associated with a disease. Corresponding assays are used for evaluation of HLA class II binding peptides. HLA class II motif-bearing peptides that are shown to bind, typically at an affinity of 1000 nM or less, are further evaluated for the ability to stimulate HTL responses.

Conventional assays utilized to detect T cell responses include proliferation assays, lymphokine secretion assays, direct cytotoxicity assays, and limiting dilution assays. For example, antigen-presenting cells that have been incubated with a peptide can be assayed for the ability to induce CTL responses in responder cell populations. Antigen-presenting cells can be normal cells such as peripheral blood mononuclear cells or dendritic cells. Alternatively, mutant non-human mammalian cell lines that are deficient in their ability to load class I molecules with internally processed peptides and that have been transfected with the appropriate human class I gene, may be used to test for the capacity of the peptide to induce *in vitro* primary CTL responses.

Peripheral blood mononuclear cells (PBMCs) may be used as the responder cell source of CTL precursors. The appropriate antigen-presenting cells are incubated with peptide, after which the peptide-loaded antigen-presenting cells are then incubated with the responder cell population under optimized culture conditions. Positive CTL activation can be determined by assaying the culture for the presence of CTLs that kill radio-labeled target cells, both specific peptide-pulsed targets as well as target cells expressing endogenously processed forms of the antigen from which the peptide sequence was derived.

More recently, a method has been devised which allows direct quantification of antigen-specific T cells by staining with Fluorescein-labelled HLA tetrameric complexes (Altman, J. D. *et al.*, *Proc. Natl. Acad. Sci. USA* 90:10330, 1993; Altman, J. D. *et al.*, *Science* 274:94, 1996). Other relatively recent technical developments include staining

for intracellular lymphokines, and interferon release assays or ELISPOT assays. Tetramer staining, intracellular lymphokine staining and ELISPOT assays all appear to be at least 10-fold more sensitive than more conventional assays (Lalvani, A. *et al.*, *J. Exp. Med.* 186:859, 1997; Dunbar, P. R. *et al.*, *Curr. Biol.* 8:413, 1998; Murali-Krishna, K. *et al.*, *Immunity* 8:177, 1998).

HTL activation may also be assessed using such techniques known to those in the art such as T cell proliferation and secretion of lymphokines, *e.g.* IL-2 (*see*, *e.g.* Alexander *et al.*, Immunity 1:751-761, 1994).

Alternatively, immunization of HLA transgenic mice can be used to determine immunogenicity of peptide epitopes. Several transgenic mouse models including mice with human A2.1, A11 (which can additionally be used to analyze HLA-A3 epitopes), and B7 alleles have been characterized and others (*e.g.*, transgenic mice for HLA-A1 and A24) are being developed. HLA-DR1 and HLA-DR3 mouse models have also been developed. Additional transgenic mouse models with other HLA alleles may be generated as necessary. Mice may be immunized with peptides emulsified in Incomplete Freund's Adjuvant and the resulting T cells tested for their capacity to recognize peptide-pulsed target cells and target cells transfected with appropriate genes. CTL responses may be analyzed using cytotoxicity assays described above. Similarly, HTL responses may be analyzed using such assays as T cell proliferation or secretion of lymphokines.

Exemplary immunogenic peptide epitopes are set out in Table XXIII.

IV.J. Use of Peptide Epitopes as Diagnostic Agents and for Evaluating Immune Responses

In one embodiment of the invention, HLA class I and class II binding peptides as described herein can be used as reagents to evaluate an immune response. The immune response to be evaluated can be induced by using as an immunogen any agent that may result in the production of antigen-specific CTLs or HTLs that recognize and bind to the peptide epitope(s) to be employed as the reagent. The peptide reagent need not be used as the immunogen. Assay systems that can be used for such an analysis include relatively recent technical developments such as tetramers, staining for intracellular lymphokines and interferon release assays, or ELISPOT assays.

For example, a peptide of the invention may be used in a tetramer staining assay to assess peripheral blood mononuclear cells for the presence of antigen-specific CTLs following exposure to a tumor cell antigen or an immunogen. The HLA-tetrameric

complex is used to directly visualize antigen-specific CTLs (*see, e.g.*, Ogg *et al.*, *Science* 279:2103-2106, 1998; and Altman *et al.*, *Science* 174:94-96, 1996) and determine the frequency of the antigen-specific CTL population in a sample of peripheral blood mononuclear cells. A tetramer reagent using a peptide of the invention may be generated as follows: A peptide that binds to an HLA molecule is refolded in the presence of the corresponding HLA heavy chain and β_2-microglobulin to generate a trimolecular complex. The complex is biotinylated at the carboxyl terminal end of the heavy chain at a site that was previously engineered into the protein. Tetramer formation is then induced by the addition of streptavidin. By means of fluorescently labeled streptavidin, the tetramer can be used to stain antigen-specific cells. The cells may then be identified, for example, by flow cytometry. Such an analysis may be used for diagnostic or prognostic purposes. Cells identified by the procedure can also be used for therapeutic purposes.

Peptides of the invention may also be used as reagents to evaluate immune recall responses. (see, *e.g.*, Bertoni *et al.*, *J. Clin. Invest.* 100:503-513, 1997 and Penna *et al.*, *J. Exp. Med.* 174:1565-1570, 1991.) For example, patient PBMC samples from individuals with HCV infection may be analyzed for the presence of antigen-specific CTLs or HTLs using specific peptides. A blood sample containing mononuclear cells may be evaluated by cultivating the PBMCs and stimulating the cells with a peptide of the invention. After an appropriate cultivation period, the expanded cell population may be analyzed, for example, for cytotoxic activity (CTL) or for HTL activity.

The peptides may also be used as reagents to evaluate the efficacy of a vaccine. PBMCs obtained from a patient vaccinated with an immunogen may be analyzed using, for example, either of the methods described above. The patient is HLA typed, and peptide epitope reagents that recognize the allele-specific molecules present in that patient are selected for the analysis. The immunogenicity of the vaccine is indicated by the presence of epitope-specific CTLs and/or HTLs in the PBMC sample.

The peptides of the invention may also be used to make antibodies, using techniques well known in the art (see, *e.g.* CURRENT PROTOCOLS IN IMMUNOLOGY, Wiley/Greene, NY; and *Antibodies A Laboratory Manual*, Harlow and Lane, Cold Spring Harbor Laboratory Press, 1989), which may be useful as reagents to diagnose or monitor cancer. Such antibodies include those that recognize a peptide in the context of an HLA molecule, *i.e.*, antibodies that bind to a peptide-MHC complex.

IV.K. Vaccine Compositions

Vaccines and methods of preparing vaccines that contain an immunogenically effective amount of one or more peptides as described herein are further embodiments of the invention. Once appropriately immunogenic epitopes have been defined, they can be sorted and delivered by various means, herein referred to as "vaccine" compositions. Such vaccine compositions can include, for example, lipopeptides (*e.g.*,Vitiello, A. *et al.*, *J. Clin. Invest.* 95:341, 1995), peptide compositions encapsulated in poly(DL-lactide-co-glycolide) ("PLG") microspheres (*see, e.g.*, Eldridge, *et al.*, *Molec. Immunol.* 28:287-294, 1991: Alonso *et al.*, *Vaccine* 12:299-306, 1994; Jones *et al.*, *Vaccine* 13:675-681, 1995), peptide compositions contained in immune stimulating complexes (ISCOMS) (*see, e.g.*, Takahashi *et al.*, *Nature* 344:873-875, 1990; Hu *et al.*, *Clin Exp Immunol.* 113:235-243, 1998), multiple antigen peptide systems (MAPs) (*see e.g.*, Tam, J. P., *Proc. Natl. Acad. Sci. U.S.A.* 85:5409-5413, 1988; Tam, J.P., *J. Immunol. Methods* 196:17-32, 1996), viral delivery vectors (Perkus, M. E. *et al.*, In: *Concepts in vaccine development*, Kaufmann, S. H. E., ed., p. 379, 1996; Chakrabarti, S. *et al.*, *Nature* 320:535, 1986; Hu, S. L. *et al.*, *Nature* 320:537, 1986; Kieny, M.-P. *et al.*, *AIDS Bio/Technology* 4:790, 1986; Top, F. H. *et al.*, *J. Infect. Dis.* 124:148, 1971; Chanda, P. K. *et al.*, *Virology* 175:535, 1990), particles of viral or synthetic origin (*e.g.*, Kofler, N. *et al.*, *J. Immunol. Methods.* 192:25, 1996; Eldridge, J. H. *et al.*, *Sem. Hematol.* 30:16, 1993; Falo, L. D., Jr. *et al.*, *Nature Med.* 7:649, 1995), adjuvants (Warren, H. S., Vogel, F. R., and Chedid, L. A. *Annu. Rev. Immunol.* 4:369, 1986; Gupta, R. K. *et al.*, *Vaccine* 11:293, 1993), liposomes (Reddy, R. *et al.*, *J. Immunol.* 148:1585, 1992; Rock, K. L., *Immunol. Today* 17:131, 1996), or, naked or particle absorbed cDNA (Ulmer, J. B. *et al.*, *Science* 259:1745, 1993; Robinson, H. L., Hunt, L. A., and Webster, R. G., *Vaccine* 11:957, 1993; Shiver, J. W. *et al.*, In: *Concepts in vaccine development*, Kaufmann, S. H. E., ed., p. 423, 1996; Cease, K. B., and Berzofsky, J. A., *Annu. Rev. Immunol.* 12:923, 1994 and Eldridge, J. H. *et al.*, *Sem. Hematol.* 30:16, 1993). Toxin-targeted delivery technologies, also known as receptor mediated targeting, such as those of Avant Immunotherapeutics, Inc. (Needham, Massachusetts) may also be used.

Vaccines of the invention include nucleic acid-mediated modalities. DNA or RNA encoding one or more of the peptides of the invention can also be administered to a patient. This approach is described, for instance, in Wolff *et. al.*, *Science* 247:1465 (1990) as well as U.S. Patent Nos. 5,580,859; 5,589,466; 5,804,566; 5,739,118; 5,736,524; 5,679,647; WO 98/04720; and in more detail below. Examples of DNA-based

delivery technologies include "naked DNA", facilitated (bupivicaine, polymers, peptide-mediated) delivery, cationic lipid complexes, and particle-mediated ("gene gun") or pressure-mediated delivery (*see, e.g.*, U.S. Patent No. 5,922,687).

For therapeutic or prophylactic immunization purposes, the peptides of the invention can also be expressed by viral or bacterial vectors. Examples of expression vectors include attenuated viral hosts, such as vaccinia or fowlpox. As an example of this approach, vaccinea virus is used as a vector to express nucleotide sequences that encode the peptides of the invention. Upon introduction into a host bearing a tumor, the recombinant vaccinia virus expresses the immunogenic peptide, and thereby elicits a host CTL and/or HTL response. Vaccinia vectors and methods useful in immunization protocols are described in, *e.g.*, U.S. Patent No. 4,722,848. Another vector is BCG (Bacille Calmette Guerin). BCG vectors are described in Stover *et al.*, *Nature* 351:456-460 (1991). A wide variety of other vectors useful for therapeutic administration or immunization of the peptides of the invention, *e.g.* adeno and adeno-associated virus vectors, retroviral vectors, *Salmonella typhi* vectors, detoxified anthrax toxin vectors, and the like, will be apparent to those skilled in the art from the description herein.

Furthermore, vaccines in accordance with the invention encompass compositions of one or more of the claimed peptide(s). A peptide can be present in a vaccine individually. Alternatively, the peptide can can exist as a homopolymer comprising multiple copies of the same peptide, or as a heteropolymer of various peptides. Polymers have the advantage of increased immunological reaction and, where different peptide epitopes are used to make up the polymer, the additional ability to induce antibodies and/or CTLs that react with different antigenic determinants of the pathogenic organism or tumor-related peptide targeted for an immune response. The composition can be a naturally occurring region of an antigen or can be prepared, *e.g.*, recombinantly or by chemical synthesis.

Carriers that can be used with vaccines of the invention are well known in the art, and include, *e.g.*, thyroglobulin, albumins such as human serum albumin, tetanus toxoid, polyamino acids such as poly L-lysine, poly L-glutamic acid, influenza, hepatitis B virus core protein, and the like. The vaccines can contain a physiologically tolerable (*i.e.*, acceptable) diluent such as water, or saline, preferably phosphate buffered saline. The vaccines also typically include an adjuvant. Adjuvants such as incomplete Freund's adjuvant, aluminum phosphate, aluminum hydroxide, or alum are examples of materials well known in the art. Additionally, as disclosed herein, CTL responses can be primed by

conjugating peptides of the invention to lipids, such as tripalmitoyl-S-glycerylcysteinlyseryl- serine (P_3CSS).

Upon immunization with a peptide composition in accordance with the invention, via injection, aerosol, oral, transdermal, transmucosal, intrapleural, intrathecal, or other suitable routes, the immune system of the host responds to the vaccine by producing large amounts of CTLs and/or HTLs specific for the desired antigen. Consequently, the host becomes at least partially immune to later infection, or at least partially resistant to developing an ongoing chronic infection, or derives at least some therapeutic benefit when the antigen was tumor-associated.

In some embodiments it may be desirable to combine the class I peptide components with components that induce or facilitate neutralizing antibody responses to the target antigen of interest, particularly to viral envelope antigens. A preferred embodiment of such a composition comprises class I and class II epitopes in accordance with the invention. An alternative embodiment of such a composition comprises a class I and/or class II epitope in accordance with the invention, along with a PADRE™ (Epimmune, San Diego, CA) molecule (described, for example, in U.S. Patent Number 5,736,142).

A vaccine of the invention can also include antigen-presenting cells, such as dendritic cells, as a vehicle to present peptides of the invention. Vaccine compositions can be created *in vitro*, following dendritic cell mobilization and harvesting, whereby loading of dendritic cells occurs *in vitro*. For example, dendritic cells are transfected, *e.g.*, with a minigene in accordance with the invention. The dendritic cell can then be administered to a patient to elicit immune responses *in vivo*.

Antigenic peptides are used to elicit a CTL and/or HTL response *ex vivo*, as well. The resulting CTL or HTL cells, can be used to treat tumors in patients that do not respond to other conventional forms of therapy, or will not respond to a therapeutic vaccine peptide or nucleic acid in accordance with the invention. *Ex vivo* CTL or HTL responses to a particular tumor-associated antigen are induced by incubating in tissue culture the patient's, or genetically compatible, CTL or HTL precursor cells together with a source of antigen-presenting cells (APC), such as dendritic cells, and the appropriate immunogenic peptide. After an appropriate incubation time (typically about 7-28 days), in which the precursor cells are activated and expanded into effector cells, the cells are infused back into the patient, where they will destroy (CTL) or facilitate destruction

(HTL) of their specific target cell (an infected cell or a tumor cell). Transfected dendritic cells may also be used as antigen presenting cells.

The vaccine compositions of the invention can also be used in combination with antiviral drugs such as interferon-α, or other treatments for viral infection.

Preferably, the following principles are utilized when selecting an array of epitopes for inclusion in a polyepitopic composition for use in a vaccine, or for selecting discrete epitopes to be included in a vaccine and/or to be encoded by nucleic acids such as a minigene. It is preferred that each of the following principles are balanced in order to make the selection. The multiple epitopes to be incorporated in a given vaccine composition may be, but need not be, contiguous in sequence in the native antigen from which the epitopes are derived.

Preferably, the following principles are utilized when selecting an array of epitopes for inclusion in a polyepitopic composition for use in a vaccine, or for selecting discrete epitopes to be included in a vaccine and/or to be encoded by nucleic acids such as a minigene. Exemplary epitopes that may be utilized in a vaccine to treat or prevent HCV infection are set out in Tables XXVI-XXIX, and Table XXXII. It is preferred that each of the following principles are balanced in order to make the selection.

1.) Epitopes are selected which, upon administration, mimic immune responses that have been observed to be correlated with HCV clearance. For HLA Class I this includes 3-4 epitopes that come from at least one antigen of HCV. For HLA Class II a similar rationale is employed; again 3-4 epitopes are selected from at least one HCV antigen (*see e.g.*, Rosenberg *et al.*, *Science* 278:1447-1450).

2.) Epitopes are selected that have the requisite binding affinity established to be correlated with immunogenicity: for HLA Class I an IC_{50} of 500 nM or less, or for Class II an IC_{50} of 1000 nM or less.

3.) Sufficient supermotif bearing-peptides, or a sufficient array of allele-specific motif-bearing peptides, are selected to give broad population coverage. For example, it is preferable to have at least 80% population coverage. A Monte Carlo analysis, a statistical evaluation known in the art, can be employed to assess the breadth, or redundancy of, population coverage.

4.) When selecting epitopes from cancer-related antigens it is often preferred to select analogs because the patient may have developed tolerance to the native epitope.

When selecting epitopes for infectious disease-related antigens it is preferable to select either native or analoged epitopes.

5.) Of particular relevance are epitopes referred to as "nested epitopes." Nested epitopes occur where at least two epitopes overlap in a given peptide sequence. A nested peptide sequence can comprise both HLA class I and HLA class II epitopes. When providing nested epitopes, it is preferable to provide a sequence that has the greatest number of epitopes per provided sequence. Preferably, one avoids providing a peptide that is any longer than the amino terminus of the amino terminal epitope and the carboxyl terminus of the carboxyl terminal epitope in the peptide. When providing a longer peptide sequence, such as a sequence comprising nested epitopes, it is important to screen the sequence in order to insure that it does not have pathological or other deleterious biological properties.

6.) If a polyepitopic protein is created, or when creating a minigene, an objective is to generate the smallest peptide that encompasses the epitopes of interest. This principle is similar, if not the same as that employed when selecting a peptide comprising nested epitopes. However, with an artificial polyepitopic peptide, the size minimization objective is balanced against the need to integrate any spacer sequences between epitopes in the polyepitopic protein. Spacer amino acid residues can be introduced to avoid junctional epitopes (an epitope recognized by the immune system, not present in the target antigen, and only created by the man-made juxtaposition of epitopes), or to facilitate cleavage between epitopes and thereby enhance epitope presentation. Junctional epitopes are generally to be avoided because the recipient may generate an immune response to that non-native epitope. Of particular concern is a junctional epitope that is a "dominant epitope." A dominant epitope may lead to such a zealous response that immune responses to other epitopes are diminished or suppressed.

Examples of polyepitopic vaccine compositions designed based on the above criteria can include epitopes from the core, S, E1, NS1/E2, NS2, NS3, NS4, and NS5 domains of the HCV polyprotein. These regions encompass the following amino acid sequences using numbering relative to the prototype HCV-1 strain (Genbank accession number M62321; *see, e.g.*, US Patent Nos. 5,683,864 and 5,670,153): C domain (amino acids 1-120); S (amino acids 120-400); NS3 (amino acids 1050-1640); NS4 (amino acids 1640-2000); NS5 (amino acids 2000-3011); and envelop proteins, E1 and E2/NS1, encompassing amino acids 192-750. Amino acids 750 to 1050 are designated as domain X as applied to the present invention. As appreciated by one of ordinary skill in the art,

the designation of the amino acid range for each domain may diverge to some extent from that of HCV-1 depending on the strain of HCV. One of ordinary skill in the art, when looking at an HCV polyprotein sequence, would readily be able to determine the domain boundaries.

Specific embodiments of the polyepitopic compositions of the present invention include a pharmaceutical composition comprising a pharmaceutically acceptable carrier and combination of motif-bearing peptides that are immunologically cross-reactive with peptides of HCV-1, wherein at least one of the peptides bears a motif of Table Ia, and further wherein the combination of motif-bearing peptides consists of: a) one or more peptides comprising at least 8 amino acids from an HCV C domain; b) one or more peptides comprising at least 8 amino acids of a further domain selected from the group consisting of: an S domain, an NS3 domain, an NS4 domain, or an NS5 domain, and; c) optionally, one or more motif-bearing peptides from one or more additional HCV domains with a *proviso* that an additional domain is not a further domain listed in "b". Preferably, such a pharmaceutical composition may additionally comprise one or more distinct HCV motif-bearing peptide(s) comprising at least 8 amino acids of an X domain or, alternatively, the composition may further comprise additional HCV motif-bearing peptide(s) that are from an envelope domain, the envelope domain peptide(s) consisting of one or more copies of a single HCV peptide comprising at least 8 amino acids of an envelope domain.

In another embodiment, the polyepitopic pharmaceutical composition may comprise a pharmaceutically acceptable carrier and combination of motif-bearing peptides that are immunologically cross-reactive with HCV-1 peptides, the peptides from multiple domains of HCV, wherein at least one of the peptides bears a motif of Table Ia, and wherein the combination of motif-bearing peptides consists essentially of: a) one or more peptides comprising at least 8 amino acids from a C domain; and, b) one or more peptides comprising at least 8 amino acids from an S, NS3, NS4, or NS5 domain, and, one HCV peptide comprising at least 8 amino acids of an envelope domain. Such a composition may further comprise one or more HCV motif-bearing peptides comprising at least 8 amino acids of an X domain.

Alternatively, a pharmaceutical composition of the invention may comprise: a) a pharmaceutically acceptable carrier; and, b) a combination of one or more motif-bearing peptides of at least 8 amino acids derived from one or more hepatitis C virus (HCV) domains, wherein said peptides are cross-reactive with peptides of HCV-1, with a *proviso*

that the combination does not include a peptide of at least 8 amino acids from an HCV C domain, and wherein at least one of the peptides bears a motif of Table Ia, said domains selected from the group consisting of: an S domain; an NS3 domain; an NS4 domain; an NS5 domain; and, an X domain. Such a composition may additionally comprise motif-bearing HCV envelope peptide(s) consisting of one or more copies of a single HCV peptide comprising at least 8 amino acids of an envelope domain.

Lastly, an embodiment of the invention may comprise a pharmaceutical composition comprising a pharmaceutically acceptable carrier and combination of two or more motif-bearing peptides from a single domain of an HCV-1 strain, said peptides immunologcially cross-reactive with peptides of an HCV-1 antigen, wherein at least one of the peptides bears a motif of Table Ia, and the peptides are derived from HCV, and the HCV domain is selected from the group consisting of: a C domain; an S domain; an NS3 domain; an NS4 domain; an NS5 domain; an X domain; or, an envelope domain from a single HCV strain, with a *proviso* that the envelope domain is other than a variable envelope domain.

In the embodiments set forth, "peptides immunologically cross-reactive with HCV-1" refers to peptides that are bound by the same antibody; "derived from" refers to a fragment or subsequence and conservatively modifed variants thereof.

IV.K.1. Minigene Vaccines

A number of different approaches are available which allow simultaneous delivery of multiple epitopes. Nucleic acids encoding the peptides of the invention are a particularly useful embodiment of the invention. Epitopes for inclusion in a minigene are preferably selected according to the guidelines set forth in the previous section. A preferred means of administering nucleic acids encoding the peptides of the invention uses minigene constructs encoding a peptide comprising one or multiple epitopes of the invention.

The use of multi-epitope minigenes is described below and in, *e.g.*, co-pending application U.S.S.N. 09/311,784; An, L. and Whitton, J. L., *J. Virol.* 71:2292, 1997; Thomson, S. A. *et al.*, *J. Immunol.* 157:822, 1996; Whitton, J. L. *et al.*, *J. Virol.* 67:348, 1993; Hanke, R. *et al.*, *Vaccine* 16:426, 1998. For example, a multi-epitope DNA plasmid encoding supermotif- and/or motif-bearing HCV epitopes derived from multiple regions of the HCV polyprotein sequence, the PADRE™ universal helper T cell epitope (or

multiple HTL epitopes from HCV), and an endoplasmic reticulum-translocating signal sequence can be engineered.

The immunogenicity of a multi-epitopic minigene can be tested in transgenic mice to evaluate the magnitude of CTL induction responses against the epitopes tested. Further, the immunogenicity of DNA-encoded epitopes *in vivo* can be correlated with the *in vitro* responses of specific CTL lines against target cells transfected with the DNA plasmid. Thus, these experiments can show that the minigene serves to both: 1.) generate a CTL response and 2.) that the induced CTLs recognized cells expressing the encoded epitopes.

For example, to create a DNA sequence encoding the selected epitopes (minigene) for expression in human cells, the amino acid sequences of the epitopes may be reverse translated. A human codon usage table can be used to guide the codon choice for each amino acid. These epitope-encoding DNA sequences may be directly adjoined, so that when translated, a continuous polypeptide sequence is created. To optimize expression and/or immunogenicity, additional elements can be incorporated into the minigene design. Examples of amino acid sequences that can be reverse translated and included in the minigene sequence include: HLA class I epitopes, HLA class II epitopes, a ubiquitination signal sequence, and/or an endoplasmic reticulum targeting signal. In addition, HLA presentation of CTL and HTL epitopes may be improved by including synthetic (*e.g.* poly-alanine) or naturally-occurring flanking sequences adjacent to the CTL or HTL epitopes; these larger peptides comprising the epitope(s) are within the scope of the invention.

The minigene sequence may be converted to DNA by assembling oligonucleotides that encode the plus and minus strands of the minigene. Overlapping oligonucleotides (30-100 bases long) may be synthesized, phosphorylated, purified and annealed under appropriate conditions using well known techniques. The ends of the oligonucleotides can be joined, for example, using T4 DNA ligase. This synthetic minigene, encoding the epitope polypeptide, can then be cloned into a desired expression vector.

Standard regulatory sequences well known to those of skill in the art are preferably included in the vector to ensure expression in the target cells. Several vector elements are desirable: a promoter with a down-stream cloning site for minigene insertion; a polyadenylation signal for efficient transcription termination; an *E. coli* origin of replication; and an *E. coli* selectable marker (*e.g.* ampicillin or kanamycin resistance). Numerous promoters can be used for this purpose, *e.g.*, the human cytomegalovirus

(hCMV) promoter. See, *e.g.*, U.S. Patent Nos. 5,580,859 and 5,589,466 for other suitable promoter sequences.

Additional vector modifications may be desired to optimize minigene expression and immunogenicity. In some cases, introns are required for efficient gene expression, and one or more synthetic or naturally-occurring introns could be incorporated into the transcribed region of the minigene. The inclusion of mRNA stabilization sequences and sequences for replication in mammalian cells may also be considered for increasing minigene expression.

Once an expression vector is selected, the minigene is cloned into the polylinker region downstream of the promoter. This plasmid is transformed into an appropriate *E. coli* strain, and DNA is prepared using standard techniques. The orientation and DNA sequence of the minigene, as well as all other elements included in the vector, are confirmed using restriction mapping and DNA sequence analysis. Bacterial cells harboring the correct plasmid can be stored as a master cell bank and a working cell bank.

In addition, immunostimulatory sequences (ISSs or CpGs) appear to play a role in the immunogenicity of DNA vaccines. These sequences may be included in the vector, outside the minigene coding sequence, if desired to enhance immunogenicity.

In some embodiments, a bi-cistronic expression vector which allows production of both the minigene-encoded epitopes and a second protein (included to enhance or decrease immunogenicity) can be used. Examples of proteins or polypeptides that could beneficially enhance the immune response if co-expressed include cytokines (*e.g.*, IL-2, IL-12, GM-CSF), cytokine-inducing molecules (*e.g.*, LeIF), costimulatory molecules, or for HTL responses, pan-DR binding proteins (PADRE™, Epimmune, San Diego, CA). Helper (HTL) epitopes can be joined to intracellular targeting signals and expressed separately from expressed CTL epitopes; this allows direction of the HTL epitopes to a cell compartment different than that of the CTL epitopes. If required, this could facilitate more efficient entry of HTL epitopes into the HLA class II pathway, thereby improving HTL induction. In contrast to HTL or CTL induction, specifically decreasing the immune response by co-expression of immunosuppressive molecules (*e.g.* TGF-β) may be beneficial in certain diseases.

Therapeutic quantities of plasmid DNA can be produced for example, by fermentation in *E. coli*, followed by purification. Aliquots from the working cell bank are used to inoculate growth medium, and grown to saturation in shaker flasks or a bioreactor

according to well known techniques. Plasmid DNA can be purified using standard bioseparation technologies such as solid phase anion-exchange resins supplied by QIAGEN, Inc. (Valencia, California). If required, supercoiled DNA can be isolated from the open circular and linear forms using gel electrophoresis or other methods.

Purified plasmid DNA can be prepared for injection using a variety of formulations. The simplest of these is reconstitution of lyophilized DNA in sterile phosphate-buffer saline (PBS). This approach, known as "naked DNA," is currently being used for intramuscular (IM) administration in clinical trials. To maximize the immunotherapeutic effects of minigene DNA vaccines, an alternative method for formulating purified plasmid DNA may be desirable. A variety of methods have been described, and new techniques may become available. Cationic lipids can also be used in the formulation (see, *e.g.*, as described by WO 93/24640; Mannino & Gould-Fogerite, *BioTechniques* 6(7): 682 (1988); U.S. Pat No. 5,279,833; WO 91/06309; and Felgner, *et al.*, *Proc. Nat'l Acad. Sci. USA* 84:7413 (1987). In addition, glycolipids, fusogenic liposomes, peptides and compounds referred to collectively as protective, interactive, non-condensing compounds (PINC) could also be complexed to purified plasmid DNA to influence variables such as stability, intramuscular dispersion, or trafficking to specific organs or cell types.

Target cell sensitization can be used as a functional assay for expression and HLA class I presentation of minigene-encoded CTL epitopes. For example, the plasmid DNA is introduced into a mammalian cell line that is suitable as a target for standard CTL chromium release assays. The transfection method used will be dependent on the final formulation. Electroporation can be used for "naked" DNA, whereas cationic lipids allow direct *in vitro* transfection. A plasmid expressing green fluorescent protein (GFP) can be co-transfected to allow enrichment of transfected cells using fluorescence activated cell sorting (FACS). These cells are then chromium-51 (^{51}Cr) labeled and used as target cells for epitope-specific CTL lines; cytolysis, detected by ^{51}Cr release, indicates both production of, and HLA presentation of, minigene-encoded CTL epitopes. Expression of HTL epitopes may be evaluated in an analogous manner using assays to assess HTL activity.

In vivo immunogenicity is a second approach for functional testing of minigene DNA formulations. Transgenic mice expressing appropriate human HLA proteins are immunized with the DNA product. The dose and route of administration are formulation dependent (*e.g.*, IM for DNA in PBS, intraperitoneal (IP) for lipid-complexed DNA).

Twenty-one days after immunization, splenocytes are harvested and restimulated for 1 week in the presence of peptides encoding each epitope being tested. Thereafter, for CTL effector cells, assays are conducted for cytolysis of peptide-loaded, ^{51}Cr-labeled target cells using standard techniques. Lysis of target cells that were sensitized by HLA loaded with peptide epitopes, corresponding to minigene-encoded epitopes, demonstrates DNA vaccine function for *in vivo* induction of CTLs. Immunogenicity of HTL epitopes is evaluated in transgenic mice in an analogous manner.

Alternatively, the nucleic acids can be administered using ballistic delivery as described, for instance, in U.S. Patent No. 5,204,253. Using this technique, particles comprised solely of DNA are administered. In a further alternative embodiment, DNA can be adhered to particles, such as gold particles.

IV.K.2. Combinations of CTL Peptides with Helper Peptides

Vaccine compositions comprising the peptides of the present invention, or analogs thereof, which have immunostimulatory activity may be modified to provide desired attributes, such as improved serum half life, or to enhance immunogenicity.

For instance, the ability of the peptides to induce CTL activity can be enhanced by linking the peptide to a sequence which contains at least one epitope that is capable of inducing a T helper cell response. The use of T helper epitopes in conjunction with CTL epitopes to enhance immunogenicity is illustrated, for example, in co-pending U.S.S.N. 08/820360, U.S.S.N. 08/197,484, and U.S.S.N. 08/464,234.

Particularly preferred CTL epitope/HTL epitope conjugates are linked by a spacer molecule. The spacer is typically comprised of relatively small, neutral molecules, such as amino acids or amino acid mimetics, which are substantially uncharged under physiological conditions. The spacers are typically selected from, *e.g.*, Ala, Gly, or other neutral spacers of nonpolar amino acids or neutral polar amino acids. It will be understood that the optionally present spacer need not be comprised of the same residues and thus may be a hetero- or homo-oligomer. When present, the spacer will usually be at least one or two residues, more usually three to six residues. Alternatively, the CTL peptide may be linked to the T helper peptide without a spacer.

Although the CTL peptide epitope can be linked directly to the T helper peptide epitope, often CTL epitope/HTL epitope conjugates are linked by a spacer molecule. The spacer is typically comprised of relatively small, neutral molecules, such as amino acids or amino acid mimetics, which are substantially uncharged under physiological

conditions. The spacers are typically selected from, *e.g.*, Ala, Gly, or other neutral spacers of nonpolar amino acids or neutral polar amino acids. It will be understood that the optionally present spacer need not be comprised of the same residues and thus may be a hetero- or homo-oligomer. When present, the spacer will usually be at least one or two residues, more usually three to six residues. The CTL peptide epitope can be linked to the T helper peptide epitope either directly or via a spacer either at the amino or carboxy terminus of the CTL peptide. The amino terminus of either the immunogenic peptide or the T helper peptide may be acylated.

HTL peptide epitopes can also be modified to alter their biological properties. For example, peptides comprising HTL epitopes can contain D-amino acids to increase their resistance to proteases and thus extend their serum half-life. Also, the epitope peptides of the invention can be conjugated to other molecules such as lipids, proteins or sugars, or any other synthetic compounds, to increase their biological activity. Specifically, the T helper peptide can be conjugated to one or more palmitic acid chains at either the amino or carboxyl termini.

In certain embodiments, the T helper peptide is one that is recognized by T helper cells present in the majority of the population. This can be accomplished by selecting amino acid sequences that bind to many, most, or all of the HLA class II molecules. These are known as "loosely HLA-restricted" or "promiscuous" T helper sequences. Examples of amino acid sequences that are promiscuous include sequences from antigens such as tetanus toxoid at positions 830-843 (QYIKANSKFIGITE), *Plasmodium falciparum* CS protein at positions 378-398 (DIEKKIAKMEKASSVFNVVNS), and Streptococcus 18kD protein at positions 116 (GAVDSILGGVATYGAA). Other examples include peptides bearing a DR 1-4-7 supermotif, or either of the DR3 motifs.

Alternatively, it is possible to prepare synthetic peptides capable of stimulating T helper lymphocytes, in a loosely HLA-restricted fashion, using amino acid sequences not found in nature (*see*, *e.g.*, PCT publication WO 95/07707). These synthetic compounds called Pan-DR-binding epitopes (*e.g.*, PADRE™, Epimmune, Inc., San Diego, CA) are designed to most preferrably bind most HLA-DR (human HLA class II) molecules. For instance, a pan-DR-binding epitope peptide having the formula: aKXVWANTLKAAa, where "X" is either cyclohexylalanine, phenylalanine, or tyrosine, and a is either D-alanine or L-alanine, has been found to bind to most HLA-DR alleles, and to stimulate the response of T helper lymphocytes from most individuals, regardless of their HLA type.

An alternative of a pan-DR binding epitope comprises all "L" natural amino acids and can be provided in the form of nucleic acids that encode the epitope.

In some embodiments it may be desirable to include in the pharmaceutical compositions of the invention at least one component which primes cytotoxic T lymphocytes. Lipids have been identified as agents capable of priming CTL *in vivo* against viral antigens. For example, palmitic acid residues can be attached to the ε-and α-amino groups of a lysine residue and then linked, *e.g.*, via one or more linking residues such as Gly, Gly-Gly-, Ser, Ser-Ser, or the like, to an immunogenic peptide. The lipidated peptide can then be administered either directly in a micelle or particle, incorporated into a liposome, or emulsified in an adjuvant, *e.g.*, incomplete Freund's adjuvant. In a preferred embodiment, a particularly effective immunogenic comprises palmitic acid attached to ε- and α- amino groups of Lys, which is attached via linkage, *e.g.*, Ser-Ser, to the amino terminus of the immunogenic peptide.

As another example of lipid priming of CTL responses, *E. coli* lipoproteins, such as tripalmitoyl-S-glycerylcysteinlyseryl- serine (P_3CSS) can be used to prime virus specific CTL when covalently attached to an appropriate peptide. (*See*, *e.g.*, Deres, *et al.*, *Nature* 342:561, 1989). Peptides of the invention can be coupled to P_3CSS, for example, and the lipopeptide administered to an individual to specifically prime a CTL response to the target antigen. Moreover, because the induction of neutralizing antibodies can also be primed with P_3CSS-conjugated epitopes, two such compositions can be combined to more effectively elicit both humoral and cell-mediated responses to infection.

As noted herein, additional amino acids can be added to the termini of a peptide to provide for ease of linking peptides one to another, for coupling to a carrier support or larger peptide, for modifying the physical or chemical properties of the peptide or oligopeptide, or the like. Amino acids such as tyrosine, cysteine, lysine, glutamic or aspartic acid, or the like, can be introduced at the C- or N-terminus of the peptide or oligopeptide, particularly class I peptides. However, it is to be noted that modification at the carboxyl terminus of a CTL epitope may, in some cases, alter binding characteristics of the peptide. In addition, the peptide or oligopeptide sequences can differ from the natural sequence by being modified by terminal-NH_2 acylation, *e.g.*, by alkanoyl (C_1-C_{20}) or thioglycolyl acetylation, terminal-carboxyl amidation, *e.g.*, ammonia, methylamine, *etc.* In some instances these modifications may provide sites for linking to a support or other molecule.

Vaccine Compositions Comprising Dendritic Cells Pulsed with CTL and/or HTL Peptides

An embodiment of a vaccine composition in accordance with the invention comprises *ex vivo* administration of a cocktail of epitope-bearing peptides to PBMC, or isolated DC therefrom, from the patient's blood. A pharmaceutical to facilitate harvesting of DC can be used, such as GM-CSF/IL-4. After pulsing the DC with peptides and prior to reinfusion into patients, the DC are washed to remove unbound peptides. In this embodiment, a vaccine comprises peptide-pulsed DCs which present the pulsed peptide epitopes complexed with HLA molecules on their surfaces. The vaccine is then administered to the patient.

IV.L. Administration of Vaccines for Therapeutic or Prophylactic Purposes

The peptides of the present invention and pharmaceutical and vaccine compositions of the invention are useful for administration to mammals, particularly humans, to treat and/or prevent HCV infection. Vaccine compositions containing the peptides of the invention are administered to a patient infected with HCV or to an individual susceptible to, or otherwise at risk for, HCV infection to elicit an immune response against HCV antigens and thus enhance the patient's own immune response capabilities. In therapeutic applications, peptide and/or nucleic acid compositions are administered to a patient in an amount sufficient to elicit an effective CTL and/or HTL response to the virus antigen and to cure or at least partially arrest or slow symptoms and/or complications. An amount adequate to accomplish this is defined as "therapeutically effective dose." Amounts effective for this use will depend on, *e.g.*, the particular composition administered, the manner of administration, the stage and severity of the disease being treated, the weight and general state of health of the patient, and the judgment of the prescribing physician.

The vaccine compositions of the invention may also be used purely as prophylactic agents. Generally the dosage for an initial prophylactic immunization generally occurs in a unit dosage range where the lower value is about 1, 5, 50, 500, or 1000 μg and the higher value is about 10,000; 20,000; 30,000; or 50,000 μg. Dosage values for a human typically range from about 500 μg to about 50,000 μg per 70 kilogram patient. This is followed by boosting dosages of between about 1.0 μg to about 50,000 μg of peptide administered at defined intervals from about four weeks to six months after the

initial administration of vaccine. The immunogenicity of the vaccine may be assessed by measuring the specific activity of CTL and HTL obtained from a sample of the patient's blood.

As noted above, peptides comprising CTL and/or HTL epitopes of the invention induce immune responses when presented by HLA molecules and contacted with a CTL or HTL specific for an epitope comprised by the peptide. The manner in which the peptide is contacted with the CTL or HTL is not critical to the invention. For instance, the peptide can be contacted with the CTL or HTL either *in vivo* or *in vitro*. If the contacting occurs *in vivo*, the peptide itself can be administered to the patient, or other vehicles, *e.g.*, DNA vectors encoding one or more peptides, viral vectors encoding the peptide(s), liposomes and the like, can be used, as described herein. When the peptide is contacted *in vitro*, the vaccinating agent can comprise a population of cells, *e.g.*, peptide-pulsed dendritic cells, or TAA-specific CTLs, which have been induced by pulsing antigen-presenting cells *in vitro* with the peptide. Such a cell population is subsequently administered to a patient in a therapeutically effective dose.

The peptides or DNA encoding them can be administered individually or as fusions of one or more peptide sequences.

For pharmaceutical compositions, the immunogenic peptides of the invention, or DNA encoding them, are generally administered to an individual already infected with HCV. The peptides or DNA encoding them can be administered individually or as fusions of one or more peptide sequences. Those in the incubation phase or the acute phase of infection can be treated with the immunogenic peptides separately or in conjunction with other treatments, as appropriate.

For therapeutic use, administration should generally begin at the first diagnosis of HCV infection. This is followed by boosting doses until at least symptoms are substantially abated and for a period thereafter. In chronic infection, loading doses followed by boosting doses may be required.

Treatment of an infected individual with the compositions of the invention may hasten resolution of the infection in acutely infected individuals. For those individuals susceptible (or predisposed) to developing chronic infection, the compositions are particularly useful in methods for preventing the evolution from acute to chronic infection. Where susceptible individuals are identified prior to or during infection, the composition can be targeted to them, thus minimizing the need for administration to a larger population.

The peptide or other compositions used for the treatment or prophylaxis of HCV infection can be used, *e.g.*, in persons who have not manifested symptoms of disease but who act as a disease vector. In this context, it is generally important to provide an amount of the peptide epitope delivered by a mode of administration sufficient to effectively stimulate a cytotoxic T cell response; compositions which stimulate helper T cell responses can also be given in accordance with this embodiment of the invention.

The dosage for an initial therapeutic immunization generally occurs in a unit dosage range where the lower value is about 1, 5, 50, 500, or 1000 μg and the higher value is about 10,000; 20,000; 30,000; or 50,000 μg. Dosage values for a human typically range from about 500 μg to about 50,000 μg per 70 kilogram patient. Boosting dosages of between about 1.0 μg to about 50000 μg of peptide pursuant to a boosting regimen over weeks to months may be administered depending upon the patient's response and condition as determined by measuring the specific activity of CTL and HTL obtained from the patient's blood. The peptides and compositions of the present invention may be employed in serious disease states, that is, life-threatening or potentially life threatening situations. In such cases, as a result of the minimal amounts of extraneous substances and the relative nontoxic nature of the peptides in preferred compositions of the invention, it is possible and may be felt desirable by the treating physician to administer substantial excesses of these peptide compositions relative to these stated dosage amounts.

Thus, for treatment of chronic infection, a representative dose is in the range disclosed above, namely where the lower value is about 1, 5, 50, 500, or 1000 μg and the higher value is about 10,000; 20,000; 30,000; or 50,000 μg, preferably from about 500 μg to about 50,000 μg per 70 kilogram patient. Initial doses followed by boosting doses at established intervals, *e.g.*, from four weeks to six months, may be required, possibly for a prolonged period of time to effectively immunize an individual. In the case of chronic infection, administration should continue until at least clinical symptoms or laboratory tests indicate that the viral infection has been eliminated or substantially abated and for a period thereafter. The dosages, routes of administration, and dose schedules are adjusted in accordance with methodologies known in the art.

The pharmaceutical compositions for therapeutic treatment are intended for parenteral, topical, oral, intrathecal, or local administration. Preferably, the pharmaceutical compositions are administered parentally, *e.g.*, intravenously,

subcutaneously, intradermally, or intramuscularly. Thus, the invention provides compositions for parenteral administration which comprise a solution of the immunogenic peptides dissolved or suspended in an acceptable carrier, preferably an aqueous carrier. A variety of aqueous carriers may be used, *e.g.*, water, buffered water, 0.8% saline, 0.3% glycine, hyaluronic acid and the like. These compositions may be sterilized by conventional, well known sterilization techniques, or may be sterile filtered. The resulting aqueous solutions may be packaged for use as is, or lyophilized, the lyophilized preparation being combined with a sterile solution prior to administration. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions, such as pH-adjusting and buffering agents, tonicity adjusting agents, wetting agents, preservatives, and the like, for example, sodium acetate, sodium lactate, sodium chloride, potassium chloride, calcium chloride, sorbitan monolaurate, triethanolamine oleate, *etc.*

The concentration of peptides of the invention in the pharmaceutical formulations can vary widely, *i.e.*, from less than about 0.1%, usually at or at least about 2% to as much as 20% to 50% or more by weight, and will be selected primarily by fluid volumes, viscosities, *etc.*, in accordance with the particular mode of administration selected.

A human unit dose form of the peptide composition is typically included in a pharmaceutical composition that comprises a human unit dose of an acceptable carrier, preferably an aqueous carrier, and is administered in a volume of fluid that is known by those of skill in the art to be used for administration of such compositions to humans (*see, e.g.*, Remington's Pharmaceutical Sciences, 17^{th} Edition, A. Gennaro, Editor, Mack Publising Co., Easton, Pennsylvania, 1985).

The peptides of the invention may also be administered via liposomes, which serve to target the peptides to a particular tissue, such as lymphoid tissue, or to target selectively to infected cells, as well as to increase the half-life of the peptide composition. Liposomes include emulsions, foams, micelles, insoluble monolayers, liquid crystals, phospholipid dispersions, lamellar layers and the like. In these preparations, the peptide to be delivered is incorporated as part of a liposome, alone or in conjunction with a molecule which binds to a receptor prevalent among lymphoid cells, such as monoclonal antibodies which bind to the CD45 antigen, or with other therapeutic or immunogenic compositions. Thus, liposomes either filled or decorated with a desired peptide of the invention can be directed to the site of lymphoid cells, where the liposomes then deliver the peptide compositions. Liposomes for use in accordance with the invention are formed

from standard vesicle-forming lipids, which generally include neutral and negatively charged phospholipids and a sterol, such as cholesterol. The selection of lipids is generally guided by consideration of, *e.g.*, liposome size, acid lability and stability of the liposomes in the blood stream. A variety of methods are available for preparing liposomes, as described in, *e.g.*, Szoka, *et al.*, *Ann. Rev. Biophys. Bioeng.* 9:467 (1980), and U.S. Patent Nos. 4,235,871, 4,501,728, 4,837,028, and 5,019,369.

For targeting cells of the immune system, a ligand to be incorporated into the liposome can include, *e.g.*, antibodies or fragments thereof specific for cell surface determinants of the desired immune system cells. A liposome suspension containing a peptide may be administered intravenously, locally, topically, *etc.* in a dose which varies according to, *inter alia*, the manner of administration, the peptide being delivered, and the stage of the disease being treated.

For solid compositions, conventional nontoxic solid carriers may be used which include, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharin, talcum, cellulose, glucose, sucrose, magnesium carbonate, and the like. For oral administration, a pharmaceutically acceptable nontoxic composition is formed by incorporating any of the normally employed excipients, such as those carriers previously listed, and generally 10-95% of active ingredient, that is, one or more peptides of the invention, and more preferably at a concentration of 25%-75%.

For aerosol administration, the immunogenic peptides are preferably supplied in finely divided form along with a surfactant and propellant. Typical percentages of peptides are 0.01%-20% by weight, preferably 1%-10%. The surfactant must, of course, be nontoxic, and preferably soluble in the propellant. Representative of such agents are the esters or partial esters of fatty acids containing from 6 to 22 carbon atoms, such as caproic, octanoic, lauric, palmitic, stearic, linoleic, linolenic, olesteric and oleic acids with an aliphatic polyhydric alcohol or its cyclic anhydride. Mixed esters, such as mixed or natural glycerides may be employed. The surfactant may constitute 0.1%-20% by weight of the composition, preferably 0.25-5%. The balance of the composition is ordinarily propellant. A carrier can also be included, as desired, as with, *e.g.*, lecithin for intranasal delivery.

IV.M. Kits

The peptide and nucleic acid compositions of this invention can be provided in kit form together with instructions for vaccine administration. Typically the kit would

include desired peptide compositions in a container, preferably in unit dosage form and instructions for administration. An alternative kit would include a minigene construct with desired nucleic acids of the invention in a container, preferably in unit dosage form together with instructions for administration. Lymphokines such as IL-2 or IL-12 may also be included in the kit. Other kit components that may also be desirable include, for example, a sterile syringe, booster dosages, and other desired excipients.

The invention will be described in greater detail by way of specific examples. The following examples are offered for illustrative purposes, and are not intended to limit the invention in any manner. Those of skill in the art will readily recognize a variety of non-critical parameters that can be changed or modified to yield alternative embodiments in accordance with the invention.

V. EXAMPLES

As in many viral diseases, there is evidence that clearance of HCV is mediated by CTL. In a study of primary HCV infection in six chimpanzees, four progressed to chronic infection (Cooper *et al.*, abstract, 19th US-Japan Hepatitis Joint Panel Meeting, January 27-29, 1998). It was found that these four animals showed either no CTL response or a very narrowly focused response during early infection. In contrast, in the remaining two animals that resolved the infection, a broad CTL response was observed against multiple HCV proteins, some of which were conserved. Weiner *et al.* (*Proc. Natl. Acad. Sci. USA* 92:2755-2759, 1995) demonstrated that viral escape, in which the epitopes presented to PATR class I molecules mutated, was linked with a progression toward chronic infection. These data show a role for the CTL in directing the course of HCV disease, and in shaping the genetic composition of HCV species in the persistently infected host.

In work in humans, Koziel and co-workers have established the presence of HCV-specific CTL in liver infiltrates from patients with chronic HCV infection (Koziel *et al.*, *J. Immunol.* 149:3339, 1992; and Koziel *et al.*, *J. Virol.* 67:7522, 1993), and have also identified a number of CTL epitopes recognized in the context of several different HLA class I molecules. Other investigators have shown that HCV-specific CTL can be detected in the peripheral blood of patients with chronic hepatitis C (Cerny *et al.*, *J. Clin. Invest.* 95:521, 1995; Cerny *et al.*, *Curr. Topics in Micro. and Immunol* 189:169, 1994; Cerny *et al.*, Abst. 2nd International Meeting on Hepatitis C and Related Viruses; La Jolla, CA, 1994; Battegay *et al.*, Abst. 2nd International Meeting on Hepatitis C and Related

Viruses; La Jolla, CA, 1994; Shirai *et al.*, *J. Virol.* 68:3334, 1994; Shirai *et al.*, *J. Immunol.* 154:2733, 1995; Battegay *et al.*, *J. Virol.* 69:2462, 1995). In addition, escape variants have been demonstrated in patients chronically infected with HCV (Chang *et al.*, *J. Clin. Invest.* 100:2376-2385, 1997; Tsai *et al.*, *Gastroenterology* 115:954-966, 1998).

The magnitude of the CTL responses observed in HCV-infected patients is, in general, higher than those observed in the case of chronic hepatitis B infection, suggesting that there is less impairment of specific T cell immunity than with HBV infection. The magnitude of CTL responses in HCV patients is, however, lower than those observed in HBV infected individuals who successfully cleared HBV infection. These results support the understanding that HCV infected patients are capable of responding to active immunotherapy, and suggest that potentiation and increasing of T cell responses to HCV may be of use in therapy and prevention of chronic HCV infection (Prince, A. M. *FEMS Micro. Rev.* 14:273, 1994).

Several groups have analyzed the potential role of HCV-specific CTL responses in disease resistance and pathogenesis. In some studies no correlation was found between CTL viremia and CTL precursor frequency for individual HCV epitopes (Rehermann *et al.*, *J. Clin. Invest.* 98:1432-1440, 1996; Wong *et al.*, *J. Immunol*. 160:1479-1488, 1998). In other studies, however, it was shown that a clear correlation existed between levels of HCV infection and CTL responses, provided that the global response against multiple CTL epitopes was considered (Rehermann *et al.*, *J. Virol.* 70:7092-7102, 1996). These data represent a strong rationale for development of vaccine constructs capable of inducing vigorous CTL responses directed against a multiplicity of conserved HCV-derived epitopes.

Koziel and colleagues have demonstrated the presence of HCV-specific CTLs, as well as T helper cell responses, in exposed but seronegative individuals (Koziel *et al.*, *J. Infect. Diseases* 176:859-866, 1997). In addition, HCV-specific CTLs have been detected in healthy, seronegative family members of chronically HCV-infected patents, indicating that a protective immunity is established in absence of a detectable infection (Bronowicki *et al.*, *J. Infect. Dis.* 176:518-522, 1997; Scognamiglio *et al.*, in preparation).

Experimental evidence also indicates that HTL epitopes play an important role in immune reactivity and defenses against HCV infection (Missale *et al.*, *J. Clin. Invest.* 98:706-714, 1996). Diepolder *et al.* (in *Lancet* 346:1006, 1995) have shown that a region of the NS3 gene (NS3 1007-1534) is recognized by patients who clear acute HCV infection, but is not seen by patients who develop chronic infection. Subsequent studies

showed that this particular region contain a highly cross-reactive HTL epitope (NS3 1248-1261), which binds with good affinity to 10 of 13 DR molecules tested, and is highly conserved in 30/33 different HCV isolates considered (Diepolder *et al., J. Virol.* 71:6011-6019, 1997). These data suggested that directing HTL responses to this type of epitope (rather than to less cross-reactive and/or highly variable ones) will be of therapeutic and prophylactic benefit and strongly argue for inclusion of this and other epitopes with similar characteristics in HCV vaccine constructs.

The following examples illustrate identification, selection, and use of immunogenic Class I and Class II peptide epitopes for inclusion in vaccine compositions.

Example 1: HLA Class I and Class II Binding Assays

The following example of peptide binding to HLA molecules demonstrates quantification of binding affinities of HLA class I and class II peptides. Binding assays can be performed with peptides that are either motif-bearing or not motif-bearing.

Epstein-Barr virus (EBV)-transformed homozygous cell lines, fibroblasts, CIR, or 721.22 transfectants were used as sources of HLA class I molecules. The specific cell lines routinely used for purification of MHC class I and class II molecules are listed in Table XXIV. Cell lysates were prepared and HLA molecules purified in accordance with disclosed protocols (Sidney *et al., Current Protocols in Immunology* 18.3.1 (1998); Sidney, *et al., J. Immunol.* 154:247 (1995); Sette, *et al., Mol. Immunol.* 31:813 (1994)). HLA molecules were purified from lysates by affinity chromatography. The lysate was passed over a column of Sepharose CL-4B beads coupled to an appropriate antibody. The antibodies used for the extraction of HLA from cell lysates are listed in Table XXV. The anti-HLA column was then washed with 10mM Tris-HCL, pH 8.0, in 1% NP-40, PBS, and PBS containing 0.4% n-octylglucoside and HLA molecules were eluted with 50mM diethylamine in 0.15M NaCl containing 0.4% n-octylglucoside, pH 11.5. A 1/25 volume of 2.0M Tris, pH 6.8, was added to the eluate to reduce the pH to ~8.0. Eluates were then be concentrated by centrifugation in Centriprep 30 concentrators (Amicon, Beverly, MA). Protein content was evaluated by a BCA protein assay (Pierce Chemical Co., Rockford, IL) and confirmed by SDS-PAGE.

A detailed description of the protocol utilized to measure the binding of peptides to Class I and Class II MHC has been published (Sette *et al., Mol. Immunol.* 31:813, 1994; Sidney *et al.*, in *Current Protocols in Immunology*, Margulies, Ed., John Wiley & Sons, New York, Section 18.3, 1998). Briefly, purified MHC molecules (5 to 500nM)

were incubated with various unlabeled peptide inhibitors and 1-10nM ^{125}I-radiolabeled probe peptides for 48h in PBS containing 0.05% Nonidet P-40 (NP40) (or 20% w/v digitonin for H-2 IA assays) in the presence of a protease inhibitor cocktail. All assays were at pH 7.0 with the exception of DRB1*0301, which was performed at pH 4.5, and DRB1*1601 (DR2w21β_1) and DRB4*0101 (DRw53), which were performed at pH 5.0.

Following incubation, MHC-peptide complexes were separated from free peptide by gel filtration on 7.8 mm x 15 cm TSK200 columns (TosoHaas 16215, Montgomeryville, PA). Because the large size of the radiolabeled peptide used for the DRB1*1501 (DR2w2β_1) assay makes separation of bound from unbound peaks more difficult under these conditions, all DRB1*1501 (DR2w2β_1) assays were performed using a 7.8mm x 30cm TSK2000 column eluted at 0.6 mls/min. The eluate from the TSK columns was passed through a Beckman 170 radioisotope detector, and radioactivity was plotted and integrated using a Hewlett-Packard 3396A integrator, and the fraction of peptide bound was determined.

Radiolabeled peptides were iodinated using the chloramine-T method. Representative radiolabeled probe peptides utilized in each assay, and its assay specific IC_{50} nM, are summarized in Tables IV and V. Typically, in preliminary experiments, each MHC preparation was titered in the presence of fixed amounts of radiolabeled peptides to determine the concentration of HLA molecules necessary to bind 10-20% of the total radioactivity. All subsequent inhibition and direct binding assays were performed using these HLA concentrations.

Since under these conditions [label]<[HLA] and IC50≥[HLA], the measured IC_{50} values are reasonable approximations of the true K_D values. Peptide inhibitors are typically tested at concentrations ranging from 120 µg/ml to 1.2 ng/ml, and are tested in two to four completely independent experiments. To allow comparison of the data obtained in different experiments, a relative binding figure is calculated for each peptide by dividing the IC_{50} of a positive control for inhibition by the IC_{50} for each tested peptide (typically unlabeled versions of the radiolabeled probe peptide). For database purposes, and inter-experiment comparisons, relative binding values are compiled. These values can subsequently be converted back into IC_{50} nM values by dividing the IC_{50} nM of the positive controls for inhibition by the relative binding of the peptide of interest. This method of data compilation has proven to be the most accurate and consistent for

comparing peptides that have been tested on different days, or with different lots of purified MHC.

Because the antibody used for HLA-DR purification (LB3.1) is α-chain specific, β_1 molecules are not separated from β_3 (and/or β_4 and β_5) molecules. The β_1 specificity of the binding assay is obvious in the cases of DRB1*0101 (DR1), DRB1*0802 (DR8w2), and DRB1*0803 (DR8w3), where no β_3 is expressed. It has also been demonstrated for DRB1*0301 (DR3) and DRB3*0101 (DR52a), DRB1*0401 (DR4w4), DRB1*0404 (DR4w14), DRB1*0405 (DR4w15), DRB1*1101 (DR5), DRB1*1201 (DR5w12), DRB1*1302 (DR6w19) and DRB1*0701 (DR7). The problem of β chain specificity for DRB1*1501 (DR2w2β_1), DRB5*0101 (DR2w2β_2), DRB1*1601 (DR2w21β_1), DRB5*0201 (DR51Dw21), and DRB4*0101 (DRw53) assays is circumvented by the use of fibroblasts. Development and validation of assays with regard to DRβ molecule specificity have been described previously (*see, e.g.*, Southwood *et al.*, *J. Immunol.* 160:3363-3373, 1998).

Binding assays as outlined above may be used to analyze supermotif and/or motif-bearing epitopes as, for example, described in Example 2.

Example 2. <u>Identification of Conserved HLA Supermotif- and Motif-Bearing CTL Candidate Epitopes</u>

Vaccine compositions of the invention may include multiple epitopes that comprise multiple HLA supermotifs or motifs to achieve broad population coverage. This example illustrates the identification of supermotif- and motif-bearing epitopes for the inclusion in such a vaccine composition. Calculation of population coverage was performed using the strategy described below.

Computer searches and algorthims for identification of supermotif and/or motif-bearing epitopes

Computer searches for epitopes bearing HLA Class I or Class II supermotifs or motifs were performed as follows. All translated HCV isolate sequences were analyzed using a text string search software program, *e.g.*, MotifSearch 1.4 (D. Brown, San Diego) to identify potential peptide sequences containing appropriate HLA binding motifs; alternative programs are readily produced in accordance with information in the art in view of the motif/supermotif disclosure herein. Furthermore, such calculations can be

made mentally. Identified A2-, A3-, and DR-supermotif sequences were scored using polynomial algorithms to predict their capacity to bind to specific HLA-Class I or Class II molecules. These polynomial algorithms take into account both extended and refined motifs (that is, to account for the impact of different amino acids at different positions), and are essentially based on the premise that the overall affinity (or ΔG) of peptide-HLA molecule interactions can be approximated as a linear polynomial function of the type:

$$\text{“}\Delta G\text{"} = a_{1i} \times a_{2i} \times a_{3i} \ldots\ldots \times a_{ni}$$

where a_{ji} is a coefficient which represents the effect of the presence of a given amino acid (j) at a given position (i) along the sequence of a peptide of n amino acids. The crucial assumption of this method is that the effects at each position are essentially independent of each other (i.e., independent binding of individual side-chains). When residue j occurs at position i in the peptide, it is assumed to contribute a constant amount j_i to the free energy of binding of the peptide irrespective of the sequence of the rest of the peptide. This assumption is justified by studies from our laboratories that demonstrated that peptides are bound to MHC and recognized by T cells in essentially an extended conformation (data omitted herein).

The method of derivation of specific algorithm coefficients has been described in Gulukota *et al.*, *J. Mol. Biol.* 267:1258-126, 1997; (see also Sidney *et al.*, *Human Immunol.* 45:79-93, 1996; and Southwood *et al.*, *J. Immunol.* 160:3363-3373, 1998). Briefly, for all i positions, anchor and non-anchor alike, the geometric mean of the average relative binding (ARB) of all peptides carrying j is calculated relative to the remainder of the group, and used as the estimate of j_i. For Class II peptides, if multiple alignments are possible, only the highest scoring alignment is utilized, following an iterative procedure. To calculate an algorithm score of a given peptide in a test set, the ARB values corresponding to the sequence of the peptide are multiplied. If this product exceeds a chosen threshold, the peptide is predicted to bind. Appropriate thresholds are chosen as a function of the degree of stringency of prediction desired.

Selection of HLA-A2 supertype cross-reactive peptides

Complete polyprotein sequences from fourteen HCV isolates were aligned, then scanned, utilizing motif identification software, to identify conserved 9- and 10-mer sequences containing the HLA-A2-supermotif main anchor specificity.

A total of 231 conserved, HLA-A2 supermotif-positive sequences were identified. These peptides were then evaluated for the presence of A*0201 preferred secondary anchor residues using A*0201-specific polynomial algorithms. A total of 67 conserved, motif-bearing and algorithm-positive sequences were identified.

Fifty of these conserved, motif-containing 9- and 10-mer peptides were tested for their capacity to bind to purified HLA-A*0201 molecules *in vitro* (HLA-A*0201 is considered a prototype A2 supertype molecule). Sixteen peptides bound A*0201 with IC_{50} values ≤500 nM; 4 with high binding affinities (IC_{50} values ≤50 nM) and 12 with intermediate binding affinities, in the 50-500 nM range (Table XXVI).

These 16 peptides were then tested for binding to additional A2-supertype molecules (A*0202, A*0203, A*0206, and A*6802). As shown in Table XXVI, most of these peptides were found to be A2-supertype cross-reactive binders. More specifically, 12/16 (75%) peptides bound at least three of the five A2-supertype molecules tested.

Selection of HLA-A3 supermotif-bearing epitopes

The sequences from the same fourteen known HCV isolates scanned above were also examined for the presence of conserved peptides with the HLA-A3-supermotif primary anchors. A total of 71 conserved 9- or 10-mer motif containing sequences were identified. Further analysis using the A03 and A11 algorithms (see, e.g., Gulukota et al, *J. Mol. Biol.* 267:1258-1267, 1997 and Sidney et al, *Human Immunol.* 45:79-93, 1996) identified 39 sequences that scored high in either or both algorithms. Twenty seven of the 39 peptides were synthesized and tested for binding to HLA-A*03 and HLA-A*11, the two most prevalent A3-supertype molecules. Fifteen peptides were identified which bound A3 and/or A11 with binding affinities of ≤500 nM (Table XXVII). These peptides were then tested for binding cross-reactivity to the other common A3-supertype alleles (A*3101, A*3301, and A*6801). Seven of the 15 peptides bound at least three of the five HLA-A3-supertype molecules tested.

In the course of an independent series of experiments (Kubo *et al.*, *J. Immunol.* 152:3913-3924, 1994), one peptide, HCV NS3 1262, not identified by the selection criteria utilized above because it does not have the A3-supermotif main anchor specificity, was determined to be cross-reactive in the A3-supertype, binding A*03, A*11, and A*6801. It is also shown in Table XXVII. Interestingly, this peptide

represents a single residue N-terminal truncation of peptide 1073.14, which is also shown in Table XXVII.

In summary, 8 peptides that bind 3 or more A3-supertype molecules derived from conserved regions of the HCV genome were identified.

Selection of HLA-B7 supermotif bearing epitopes

When the same fourteen HCV isolates were also analyzed for the presence of conserved 9- or 10-mer peptides with the HLA-B7-supermotif, 35 sequences were identified. The corresponding peptides were synthesized and tested for binding to HLA-B*0702, the most common B7-supertype allele (*i.e.*, the prototype B7 supertype allele). Thirteen peptides bound B*0702 with IC_{50} of ≤500 nM (Table XXVIIIa). These 13 peptides were then tested for binding to other common B7-supertype molecules (B*3501, B*51, B*5301, and B*5401). As shown in Table XXVIIIa, only 1 peptide (Core 169) was capable of binding to three or more of the five B7-supertype alleles tested.

To identify additional B7-supertype epitopes, further studies were undertaken. The protein sequences from the fourteen HCV isolates utilized above were again examined to identify conserved, motif-containing 8- and 11-mers. The isolates were also examined for 9- and 10-mer sequences allowing for lower conservancy (51%-78%). Twenty-five 8-mers, sixteen 11-mers, and thirty-five 9- and 10-mers were identified, synthesized, and tested for binding to B*0702. Thirteen peptides bound with high or intermediate affinity (IC_{50} ≤500 nM) (Table XXVIIIb). These peptides were additionally screened for binding to other B7-supertype molecules. Only one cross-reactive binder, the NS3 1378 8-mer (peptide 29.0035/1260.04), was identified (Table XXVIIIb).

In summary, a total of two cross-reactive B7-supertype binders were identified (Core 169 and NS3 1378).

Selection of A1 and A24 motif-bearing epitopes

To further increase population coverage, HLA-A1 and -A24 epitopes can also be incorporated into potential vaccine constructs.

In a previous analysis, two A1 and three A24 binders, 100% conserved among four strains of HCV, were identified (Wentworth *et al.*, *Int. Immunol.* 8:651-659, 1996). An analysis of the protein sequence data from the fourteen HCV strains utilized above demonstrated that these peptides were >79% conserved, and also identified an additional

eleven A1- and twenty five A24-motif-containing conserved sequences (see Table XXIXA and B). Eight of the additional eleven A1 peptides and seven of the additional twenty five A24 peptides were tested for binding to the appropriate HLA molecule (*i.e.*, A1 or A24). Overall, as shown in Table XXIX, four A1-motif peptides (A) and three A24-motif peptides (B) have been found with binding capacities of 500 nM or less for the appropriate allele-specific HLA molecule.

Analysis of the HLA-A2 and A3 supermotif-bearing epitopes identified above revealed that in 13/14 cases, peptides binding the supertype prototype HLA molecule (i.e. A*0201 for the A2 supertype, and A*0301 for the A3 supertype) with an IC_{50} of less than 100nM were cross-reactive and recognized by HCV-infected patients as described in Example 3, which follows. Based on these observations, two A1 peptides and one A24 peptide epitopes were also selected as candidates for inclusion in vaccine compositions; these peptides bind the appropriate HLA molecule with an IC_{50} of less than 100nM.

Example 3: Confirmation of Immunogenicity

*Evaluation of A*0201 immunogenicity*

It has been shown that CTL induced in A*0201/K^b transgenic mice exhibit specificity similar to CTL induced in the human system (*see, e.g.*, Vitiello *et al.*, *J. Exp. Med.* 173:1007-1015, 1991; Wentworth *et al.*, *Eur. J. Immunol.* 26:97-101, 1996). Accordingly, these mice were used to evaluate the immunogenicity of the twelve conserved A2-supertype cross-reactive peptides identified in Example 2 above.

CTL induction in transgenic mice following peptide immmunization has been described (Vitiello *et al.*, *J. Exp. Med.* 173:1007-1015, 1991; Alexander *et al.*; *J. Immunol.* 159:4753-4761, 1997). In these studies, mice were injected subcutaneously at the base of the tail with each peptide (50 µg/mouse) emulsified in IFA in the presence of an excess of an IA^b-restricted helper peptide (140 µg/mouse) (HBV core 128-140, Sette *et al.*, *J. Immunol.* 153:5586-5592, 1994). Eleven days after injection, splenocytes were incubated in the presence of peptide-loaded syngenic LPS blasts. After six days, cultures were assayed for cytotoxic activity using peptide-pulsed targets. The data, summarized in Table XXX, indicate that 7 of the 12 peptides (58%) were capable of inducing primary CTL responses in A*0201/K^b transgenic mice. (For these studies, a peptide was considered positive if it induced CTL (L.U. 30/10^6 cells ≥2 in at least two transgenic animals (Wentworth *et al.*, *Eur. J. Immunol.* 26:97-101, 1996).

The conserved, cross reactive candidate CTL epitopes were also tested for recognition *in vitro* by PBMCs obtained from HCV-infected patients. Briefly, PBMC from patients infected with HCV were cultured in the presence of 10 μg/ml of synthetic peptide. After 7 and 14 days, the cultures were restimulated with peptide. The cultures were assayed for cytolytic activity on day 21 using target cells pulsed with the specific peptide in a standard four hour ^{51}Cr release assay. The data are summarized in Table XXX. As shown, all 12 peptides are CTL epitopes recognized by PBMC from HCV-infected patients. From the data in Table XXX, it is interesting to note that HLA transgenics did not fully reveal the immunogenicity of some peptides that were positive in recall responses. This apparent discrepancy may reflect differences in the route of immunization utilized (*e.g.*, natural infection versus peptide immunization), or CTL repertoire.

*Evaluation of A*03/A11 immunogenicity*

The immunogenicity of six of the eight A3-supertype cross-reactive peptides identified in Example 2 above was evaluated in HLA-A11/K^b transgenic mice, using the protocol described above for HLA-A2 transgenic mice (Alexander *et al.*, *J. Immunol.* 159:4753-4761, 1997). Five of these six peptides were able to induce primary CTL responses (Table XXXI).

All eight peptides were also studied by collaborators using PBMC cultures from HCV infected patients and contacts of such patients. This data is also summarized in Table XXXI. Briefly, all eight peptides were recognized by HCV infected individuals.

Evaluation of B7 immunogenicity

One of the two B7-supertype cross-reactive peptides (1145.12, Core 169) has been evaluated for immunogenicity in HCV-infected patients. Two independent collaborators have shown that this peptide is indeed immunogenic, and is recognized by T cells from HCV-infected patients (Chang *et al.*, *J. Immunol.* 162:1156-1164, 1999)

Example 4: Implementation of the Extended Supermotif to Improve the Binding Capacity of Native Epitopes by Creating Analogs

HLA motifs and supermotifs (comprising primary and/or secondary residues) are useful in the identification and preparation of highly cross-reactive native peptides, as demonstrated herein. Moreover, the definition of HLA motifs and supermotifs also

allows one to engineer highly cross-reactive epitopes by identifying residues within a native peptide sequence which can be analogued, or "fixed" to confer upon the peptide certain characteristics, *e.g.* greater cross-reactivity within the group of HLA molecules that comprise a supertype, and/or greater binding affinity for some or all of those HLA molecules. Examples of analog peptides that exhibit modulated binding affinity are set forth in this example.

Analoging at Primary Anchor Residues

As shown in Example 2, more than ten different HCV-derived, A2-supertype-restricted epitopes were identified. Peptide engineering strategies are implemented to further increase the cross-reactivity of the candidate epitopes identified above which bind 3/5 of the A2 supertype alleles tested. On the basis of the data disclosed, *e.g.*, in related and co-pending U.S.S.N 09/226,775, the main anchors of A2-supermotif-bearing peptides are altered, for example, to introduce a preferred L, I, V, or M at position 2, and I or V at the C-terminus.

To analyze the cross-reactivity of the analog peptides, each engineered analog is initially tested for binding to the prototype A2 supertype allele A*0201, then, if A*0201 binding capacity is maintained, for A2-supertype cross-reactivity.

Similarly, analogs of HLA-A3 supermotif-bearing epitopes may also be generated. For example, peptides binding to 3/5 of the A3-supertype molecules may be engineered at primary anchor residues to possess a preferred residue (V, S, M, or A) at position 2.

The analog peptides are then tested for the ability to bind A*03 and A*11 (prototype A3 supertype alleles). Those peptides that demonstrate ≤ 500 nM binding capacity are then tested for A3-supertype cross-reactivity.

Similarly to the A2- and A3- motif bearing peptides, peptides binding 3 or more B7-supertype alleles may be improved, where possible, to achieve increased cross-reactive binding. B7 supermotif-bearing peptides may, for example, be engineered to possess a preferred residue (V, I, L, or F) at the C-terminal primary anchor position, as demonstrated by Sidney *et al.* (*J. Immunol.* 157:3480-3490, 1996).

Analoging at Secondary Anchor Residues

Moreover, HLA supermotifs are of value in engineering highly cross-reactive peptides and/or peptides that bind HLA molecules with increased affinity by identifying

particular residues at secondary anchor positions that are associated with such properties. Demonstrating this, the binding capacity of a peptide representing a discreet single amino acid substitution at position one was analyzed. Peptide 1145.13 (Table XXVIIIc), which represents the substitution of L to F at position 1 of the core 169 sequence, binds all five B7-supertype molecules with a good affinity (all IC_{50} values ≤ 132 nM), and in 3 instances has higher affinity over that of the parent peptide by >35-fold.

Because so few B7-supertype cross-reactive epitopes were identified, our results from previous binding evaluations were analyzed to identify conserved (8-, 9-, 10-, or 11-mer) peptides which bind, minimally, 3/5 B7 supertype molecules with weak affinity (IC_{50} of 500nM-5μM). This analysis identified 9 peptides, 6 of which are analogued (including core 169 which had been previously analogued). These peptides are tested for enhanced binding affinity and B7-supertype cross-reactivity.

Engineered analogs with sufficiently improved binding capacity or cross-reactivity are tested for immunogenicity in HLA-B7-transgenic mice, following for example, IFA immunization or lipopeptide immunization.

In conclusion, these data demonstrate that by the use of even single amino acid substitutions, it is possible to increase the binding affinity and/or cross-reactivity of peptide ligands for HLA supertype molecules.

Example 5: <u>Identification of conserved HCV-derived sequences with HLA-DR binding motifs</u>

Peptide epitopes bearing an HLA class II supermotif or motif may also be identified as outlined below using methodology similar to that described in Examples 1-3.

Selection of HLA-DR-supermotif-bearing epitopes

To identify HCV-derived, HLA class II HTL epitopes, the same fourteen HCV polyprotein sequences used for the identification of HLA Class I supermotif/motif sequences were analyzed for the presence of sequences bearing an HLA-DR-motif or supermotif. Specifically, 15-mer sequences were selected comprising a DR-supermotif, further comprising a 9-mer core, and three-residue N- and C-terminal flanking regions (15 amino acids total). It was also required that the 15-mer sequence be conserved in at least 79% (11/14) of the HCV strains analyzed. These criteria identified a total of 49 non-redundant sequences, which are shown in Table XXXIIA. (In the context of Class II

epitopes, a sequence is considered operationally redundant if more than 80% of it's sequence overlaps with another peptide.)

Protocols for predicting peptide binding to DR molecules have been developed (Southwood *et al., J. Immunol.* 160:3363-3373, 1998). These protocols, specific for individual DR molecules, allow the scoring, and ranking, of 9-mer core regions. Each protocol not only scores peptide sequences for the presence of DR-supermotif primary anchors (i.e., at position 1 and position 6) within a 9-mer core, but additionally evaluates sequences for the presence of secondary anchors. Using allele specific selection tables (see, *e.g.*, Southwood *et al.*, *ibid.*), it has been found that these protocols efficiently select peptide sequences with a high probability of binding a particular DR molecule. Additionally, it has been found that performing these protocols in tandem, specifically those for DR1, DR4w4, and DR7, can efficiently select DR cross-reactive peptides.

To see if these protocols serve to identify additional epitopes, the same HCV polyproteins used above were re-scanned for the presence of 15-mer peptides with 9-mer core regions that were ≥79% (11/14 strains) conserved. This identified 152 sequences; 49 of which were identified previously, as described above. Next, the 9-mer core region of each of these peptides was scored using the DR1, DR4w4, and DR7 algorithms. Twenty-two peptides, including 12 new sequences (10 peptides were from the original set of 49) were found to have 9-mer cores with protocol-derived scores predictive of cross-reactive DR binders. The 12 additional sequences are shown in Table XXXIIB.

The conserved, HCV-derived peptides identified above were tested for their binding capacity for various common HLA-DR molecules. All peptides were initially tested for binding to the DR molecules in the primary panel: DR1, DR4w4, and DR7. Peptides binding at least 2 of these 3 DR molecules were then tested for binding to DR2w2 β1, DR2w2 β2, DR6w19, and DR9 molecules in secondary assays. Finally, peptides binding at least 2 of the 4 secondary panel DR molecules, and thus cumulatively at least 4 of 7 different DR molecules, were screened for binding to DR4w15, DR5w11, and DR8w2 molecules in tertiary assays. Peptides binding at least 7 of the 10 DR molecules comprising the primary, secondary, and tertiary screening assays were considered cross-reactive DR binders. The composition of these screening panels, and the phenotypic frequency of associated antigens, are shown in Table XXXIII.

Upon testing, it was found that 29 of the original 75 peptides (39%) bound two or more of the primary HLA molecules. Twenty-six of these cross-reactive binders were

then tested in the secondary assays, and nineteen were found to bind at least four of the seven HLA DR molecules in the primary and secondary panels. Finally, the nineteen peptides passing the secondary screening phase were tested for binding in the tertiary assays. As a result, nine peptides were identified which bound at least seven of ten common HLA-DR molecules. Table XXXIV shows these nine peptides and their binding capacity for each allele-specific HLA-DR molecule in the primary through tertiary panels. Also shown in Table XXXIV are two peptides (F134.05 and F134.08) for which a complete binding analysis was not performed. However, both of these peptides bound six of the seven HLA DR molecules tested. F134.08 nests peptide 1283.44, which bound eight of 10 allele-specific HLA molecules.

In conclusion, eleven cross-reactive DR-binding peptides, derived from six discrete (*i.e.* non-redundant) regions of the HCV genome, have been identified. Two of the six regions from which these epitopes were derived are covered by multiple, overlapping epitopes.

Selection of conserved DR3 motif peptides

Because HLA-DR3 is an allele that is prevalent in Caucasian, Black, and Hispanic populations, DR3 binding capacity is an important criterion in the selection of HTL epitopes. However, data generated previously indicated that DR3 only rarely cross-reacts with other DR alleles (Sidney *et al., J. Immunol.* 149:2634-2640, 1992; Geluk *et al., J. Immunol.* 152:5742-5748, 1994; Southwood *et al., J. Immunol.* 160:3363-3373, 1998). This is not entirely surprising in that the DR3 peptide-binding motif appears to be distinct from the specificity of most other DR alleles.

To efficiently identify peptides that bind DR3, target proteins were analyzed for conserved sequences carrying one of the two DR3 specific binding motifs reported by Geluk *et al.* (*J. Immunol.* 152:5742-5748, 1994). Fifteen sequences, including a peptide nested within a DR-supermotif sequence identified above (peptide Pape 22), were identified (Table XXXIId). Preferably, DR3 motifs will be found clustered in proximity with DR supermotif regions.

Fourteen of the fifteen peptides containing a DR3 motif were tested for their DR3 binding capacity. Two peptides (CH35.0106 and CH35.0107) were found to bind DR3 with an affinity of 1μM or less (Table XXXV), and thereby qualify as HLA class II high affinity binders.

DR3 binding epitopes identified in this manner may then be included in vaccine compositions with DR supermotif-bearing peptide epitopes.

Example 6: Immunogenicity of candidate HCV-derived HTL epitopes and known dominant HCV HTL epitope

In the course of collaborative studies with G. Pape and C. Ferrari, eight conserved, HCV-derived peptides have been identified which are recognized by HCV-infected individuals.

One of these studies (Diepolder *et al.*, *J. Virol.* 71:6011-6019, 1997), identified peptide F98.05, which spans residues 1248-1261 of the NS3 protein, as an immunodominant CD4+ T-cell epitope that was recognized by 14/23 NS3-specific CD4+ T-cell clones from 4/5 patients with acute hepatitis C infection. This epitope, shown above to be an HLA-DR cross-reactive binder (see Table XXXIV), was capable of being presented to helper CD4+ T cells by multiple HLA molecules (DR4, DR11, DR12, DR13, and DR16). Two other peptides, Pape 22 and Pape 29, were also recognized by CD4+ T cell clones, although, in a more limited context; correspondingly, neither of these peptides are DR-cross-reactive binders.

By direct peripheral blood T cell stimulation and by fine specificity analysis of HCV-specific T-cell lines and clones, studies done in collaboration with Ferrari's group identified 6 immunodominant epitopes, including one also identified in the Pape collaboration, that are derived from conserved regions of the core, NS3, and NS4 proteins. These epitopes were also found to be cross-reactive, being presented to T cells in the context of different Class II molecules. Three of the 6 epitopes, F98.04 (F134.03), F134.05 and F134.08, are cross-reactive HLA-DR binders (see Table XXXIV).

In conclusion, the immunogenicity of 8 epitopes derived from conserved regions of the HCV genome has been demonstrated. Three of these epitopes (F98.05, F134.05, and F134.08; see Table XXXIV) are broadly cross-reactive HLA-DR binding peptides.

Example 7. Calculation of phenotypic frequencies of HLA-supertypes in various ethnic backgrounds to determine breadth of population coverage

This example illustrates the assessment of the breadth of population coverage of a vaccine composition comprised of multiple epitopes comprising multiple supermotifs and/or motifs.

In order to analyze population coverage, gene frequencies of HLA alleles were determined. Gene frequencies for each HLA allele were calculated from antigen or allele frequencies utilizing the binomial distribution formulae gf=1-(SQRT(1-af)) (see, *e.g.*, Sidney *et al.*, *Human Immunol.* 45:79-93, 1996). To obtain overall phenotypic frequencies, cumulative gene frequencies were calculated, and the cumulative antigen frequencies derived by the use of the inverse formula [af=1-(1-Cgf)^2].

Where frequency data was not available at the level of DNA typing, correspondence to the serologically defined antigen frequencies was assumed. To obtain total potential supertype population coverage no linkage disequilibrium was assumed, and only alleles confirmed to belong to each of the supertypes were included (minimal estimates). Estimates of total potential coverage achieved by inter-loci combinations were made by adding to the A coverage the proportion of the non-A covered population that could be expected to be covered by the B alleles considered (*e.g.*, total=A+B*(1-A)). Confirmed members of the A3-like supertype are A3, A11, A31, A*3301, and A*6801. Although the A3-like supertype may also include A34, A66, and A*7401, these alleles were not included in overall frequency calculations. Likewise, confirmed members of the A2-like supertype family are A*0201, A*0202, A*0203, A*0204, A*0205, A*0206, A*0207, A*6802, and A*6901. Finally, the B7-like supertype-confirmed alleles are: B7, B*3501-03, B51, B*5301, B*5401, B*5501-2, B*5601, B*6701, and B*7801 (potentially also B*1401, B*3504-06, B*4201, and B*5602).

Population coverage achieved by combining the A2-, A3- and B7-supertypes is approximately 86% in five major ethnic groups (see Table XXI). Coverage may be extended by including peptides bearing the A1 and A24 motifs. On average, A1 is present in 12% and A24 in 29% of the population across five different major ethnic groups (Caucasian, North American Black, Chinese, Japanese, and Hispanic). Together, these alleles are represented with an average frequency of 39% in these same ethnic populations. The total coverage across the major ethnicities when A1 and A24 are combined with the coverage of the A2-, A3- and B7-supertype alleles is >95%. An analagous approach can be used to estimate population coverage achieved with combinations of class II motif-bearing epitopes.

Summary of candidate HLA class I and class II epitopes

In summary, on the basis of the data presented in the above examples, 26 CTL candidate peptide epitopes derived from conserved regions of the HCV virus have been

identified (Table XXXVIa). These include twelve HLA-A2 supermotif-bearing epitopes, eight HLA-A3 supermotif-bearing epitopes, and one HLA-B7 supermotif-bearing epitope, each capable of binding to multiple A2-, A3-, or B7-supertype molecules, and immunogenic in HLA transgenic mice or antigenic for human PBL (with the exception of peptide 29.0035/1260.04). Additional epitopes not evaluated for immunogenicity are also included. They are an additional B7-supermotif-bearing epitope and two HLA-A1 and one HLA-A24 high-affinity binding peptides. A known HLA-A31 restricted epitope (VGIYLLPNR), which also binds HLA-A33, is also set out in Table XXXVIa and is useful in combination with other Class I or Class II epitopes.

With these 26 CTL epitopes (as disclosed herein and from the art), average population coverage, (*i.e.*, recognition of at least one HCV epitope), is predicted to be greater than 95% in each of five major ethnic populations. The potential redundancy of coverage afforded by 25 of these epitopes (the peptide 24.0086 was not included) was estimated using the game theory Monte Carlo simulation analysis, which is known in the art (see *e.g.*, Osborne, M.J. and Rubinstein, A. "A course in game theory" MIT Press, 1994). As shown in Figure 1, it is estimated that 90% of the individuals in a population comprised of the Caucasian, North American Black, Japanese, Chinese, and Hispanic ethnic groups would recognize 2 or more of the candidate epitopes described herein.

A list of HCV-derived HTL epitopes that would be preferred for use in the design of minigene constructs or other vaccine formulations is summarized in Table XXXVIb. As shown, 9 different peptide-binding regions have been identified which bind multiple HLA-DR molecules or bind HLA-DR3. (In the case of the NS4 1914-1935 region, the longer peptide, F134.08, recognized by patients, was chosen over the shorter peptide, 1283.44. The longer peptide essentially incorporates the shorter peptide, and also binds additional DR molecules that the shorter peptide does not bind.) Three of these peptides have been recognized as dominant epitopes in HCV infected patients.

It is estimated that each of 10 common DR molecules recognizing the DR supermotif, and DR3, are covered by a minimum of 2 epitopes. Correspondingly, the total estimated population coverage represented by this panel of epitopes is in excess of 91% in each of the 5 major ethnic populations (Table XXXVII).

Example 8: Recognition Of Generation Of Endogenous Processed Antigens After Priming

This example determines that CTL induced by native or analogued peptide epitopes identified and selected as described in Examples 1-6 recognize endogenously synthesized, *i.e.*, native antigens.

Effector cells isolated from transgenic mice that are immunized with peptide epitopes as in Example 3, for example HLA-A2 supermotif-bearing epitopes, are re-stimulated *in vitro* using peptide-coated stimulator cells. Six days later, effector cells are assayed for cytotoxicity and the cell lines that contain peptide-specific cytotoxic activity are further re-stimulated. An additional six days later, these cell lines are tested for cytotoxic activity on ^{51}Cr labeled Jurkat-A2.1/K^{b} target cells in the absence or presence of peptide, and also tested on ^{51}Cr labeled target cells bearing the endogenously synthesized antigen, *i.e.* cells that are stably transfected with HCV expression vectors.

The result will demonstrate that CTL lines obtained from animals primed with peptide epitope recognize endogenously synthesized HCV antigen. The choice of transgenic mouse model to be used for such an analysis depends upon the epitope(s) that is being evaluated. In addition to HLA-A*0201/K^{b} transgenic mice, several other transgenic mouse models including mice with human A11, which may also be used to evaluate A3 epitopes, and B7 alleles have been characterized and others (*e.g.*, transgenic mice for HLA-A1 and A24) are being developed. HLA-DR1 and HLA-DR3 mouse models have also been developed, which may be used to evaluate HTL epitopes.

Example 9: Activity Of CTL-HTL Conjugated Epitopes In Transgenic Mice

This example illustrates the induction of CTLs and HTLs in transgenic mice by use of an HCV CTL/HTL peptide conjugate whereby the vaccine composition comprises peptides administered to an HCV-infected patient or an individual at risk for HCV. The peptide composition can comprise multiple CTL and/or HTL epitopes. This analysis demonstrates enhanced immunogenicity that can be achieved by inclusion of one or more HTL epitopes in a vaccine composition. Such a peptide composition can comprise a lipidated HTL epitope conjugated to a preferred CTL epitope containing, for example, at least one CTL epitope selected from Table XXVI-XXIX, or an analog of that epitope. The HTL epitope is, for example, selected from Table XXXII.

Lipopeptide preparation: Lipopeptides are prepared by coupling the appropriate fatty acid to the amino terminus of the resin bound peptide. A typical procedure is as

follows: A dichloromethane solution of a four-fold excess of a pre-formed symmetrical anhydride of the appropriate fatty acid is added to the resin and the mixture is allowed to react for two hours. The resin is washed with dichloromethane and dried. The resin is then treated with trifluoroacetic acid in the presence of appropriate scavengers [e.g. 5% (v/v) water] for 60 minutes at 20°C. After evaporation of excess trifluoroacetic acid, the crude peptide is washed with diethyl ether, dissolved in methanol and precipitated by the addition of water. The peptide is collected by filtration and dried.

Immunization procedures: Immunization of transgenic mice is performed as described (Alexander *et al., J. Immunol.* 159:4753-4761, 1997). For example, A2/K^b mice, which are transgenic for the human HLA A2.1 allele and are useful for the assessment of the immunogenicity of HLA-A*0201 motif- or HLA-A2 supermotif-bearing epitopes, are primed subcutaneously (base of the tail) with 0.1 ml of peptide conjugate formulated in saline, or DMSO/saline. Seven days after priming, splenocytes obtained from these animals are restimulated with syngenic irradiated LPS-activated lymphoblasts coated with peptide.

Cell lines: Target cells for peptide-specific cytotoxicity assays are Jurkat cells transfected with the HLA-A2.1/K^b chimeric gene (*e.g.*, Vitiello *et al., J. Exp. Med.* 173:1007, 1991)

In vitro CTL activation: One week after priming, spleen cells ($30x10^6$ cells/flask) are co-cultured at 37°C with syngeneic, irradiated (3000 rads), peptide coated lymphoblasts ($10x10^6$ cells/flask) in 10 ml of culture medium/T25 flask. After six days, effector cells are harvested and assayed for cytotoxic activity.

Assay for cytotoxic activity: Target cells (1.0 to $1.5x10^6$) are incubated at 37°C in the presence of 200 µl of ^{51}Cr. After 60 minutes, cells are washed three times and resuspended in R10 medium. Peptide is added where required at a concentration of 1 µg/ml. For the assay, 10^4 ^{51}Cr-labeled target cells are added to different concentrations of effector cells (final volume of 200 µl) in U-bottom 96-well plates. After a 6 hour incubation period at 37°C, a 0.1 ml aliquot of supernatant is removed from each well and radioactivity is determined in a Micromedic automatic gamma counter. The percent specific lysis is determined by the formula: percent specific release = 100 x (experimental release - spontaneous release)/(maximum release - spontaneous release). To facilitate comparison between separate CTL assays run under the same conditions, % ^{51}Cr release data is expressed as lytic units/10^6 cells. One lytic unit is arbitrarily defined as the number of effector cells required to achieve 30% lysis of 10,000 target cells in a 6

hour ^{51}Cr release assay. To obtain specific lytic units/10^6, the lytic units/10^6 obtained in the absence of peptide is subtracted from the lytic units/10^6 obtained in the presence of peptide. For example, if 30% ^{51}Cr release is obtained at the effector (E): target (T) ratio of 50:1 (i.e., $5x10^5$ effector cells for 10,000 targets) in the absence of peptide and 5:1 (i.e., $5x10^4$ effector cells for 10,000 targets) in the presence of peptide, the specific lytic units would be: $[(1/50{,}000)-(1/500{,}000)] \times 10^6 = 18$ LU.

The results are analyzed to assess the magnitude of the CTL responses of animals injected with the immunogenic CTL/HTL conjugate vaccine preparation and are compared to the magnitude of the CTL response achieved using the CTL epitope as outlined in Example 3. Analyses similar to this may be performed to evaluate the immunogenicity of peptide conjugates containing multiple CTL epitopes and/or multiple HTL epitopes. In accordance with these procedures it is found that a CTL response is induced, and concomitantly that an HTL response is induced upon administration of such compositions.

Example 10. Selection of CTL and HTL epitopes for inclusion in an HCV-specific vaccine.

This example illustrates the procedure for the selection of peptide epitopes for vaccine compositions of the invention. The peptides in the composition can be in the form of a nucleic acid sequence, either single or one or more sequences (*i.e.*, minigene) that encodes peptide(s), or may be single and/or polyepitopic peptides.

Epitopes are selected which, upon administration, mimic immune responses that have been observed to be correlated with tumor clearance. For example, vaccine can include 3-4 epitopes that come from at least one HCV antigen region. Epitopes from one region can be used in combination with epitopes from one or more additional HCV antigen regions. Analogs of epitopes can also be selected for inclusion in the vaccine.

Epitopes are often selected that have a binding affinity of an IC_{50} of 500 nM or less for an HLA class I molecule, or for class II, an IC_{50} of 1000 nM or less.

Sufficient supermotif bearing peptides, or a sufficient array of allele-specific motif bearing peptides, are selected to give broad population coverage. For example, epitopes are selected to provide at least 80% population coverage. A Monte Carlo analysis, a statistical evaluation known in the art, can be employed to assess breadth, or redundancy, of population coverage.

When creating a polyepitopic compositions, *e.g.* a minigene, it is typically desirable to generate the smallest peptide possible that encompasses the epitopes of interest. The principles employed are similar, if not the same, as those employed when selecting a peptide comprising nested epitopes. Additionally, however, upon determination of the nucleic acid sequence to be provided as a minigene, the peptide sequence encoded thereby is analyzed to determine whether any "junctional epitopes" have been created. A junctional epitope is a potential HLA binding epitope, as predicted, *e.g.*, by motif analysis. Junctional epitopes are generally to be avoided because the recipient may bind to an HLA molecule and generate an immune response to that epitope, which is not present in a native protein sequence.

Peptide epitopes for inclusion in vaccine compositions are, for example, selected from those listed in Tables XXVI-XXIX and Table XXXII. A vaccine composition comprised of selected peptides, when administered, is safe, efficacious, and elicits an immune response similar in magnitude of an immune response that clears an acute HCV infection.

Example 11: Construction of Minigene Multi-Epitope DNA Plasmids

This example provides guidance for the construction of a minigene expression plasmid. Minigene plasmids may, of course, contain various configurations of CTL and/or HTL epitopes or epitope analogs as described herein. Examples of the construction and evaluation of expression plasmids are described, for example, in co-pending U.S.S.N. 09/311,784 filed 5/13/99. An example of such a plasmid for the expression of HCV epitopes is shown in Figure 2, which illustrates the orientation of HCV peptide epitopes in a minigene construct.

A minigene expression plasmid may include multiple CTL and HTL peptide epitopes. In the present example, HLA-A2, -A3, -B7 supermotif-bearing peptide epitopes and HLA-A1 and -A24 motif-bearing peptide epitopes are used in conjunction with DR supermotif-bearing epitopes and/or DR3 epitopes (Figure 2). Preferred epitopes are identified, for example, in Tables XXVI-XXIX and XXXII. HLA class I supermotif or motif-bearing peptide epitopes derived from multiple HCV antigens, *e.g.*, the core, NS4, NS3, NS5, NS1/E2, are selected such that multiple supermotifs/motifs are represented to ensure broad population coverage. Similarly, HLA class II epitopes are selected from multiple HCV antigens to provide broad population coverage, *i.e.* both HLA DR-1-4-7 supermotif-bearing epitopes and HLA DR-3 motif-bearing epitopes are selected for

inclusion in the minigene construct. The selected CTL and HTL epitopes are then incorporated into a minigene for expression in an expression vector.

This example illustrates the methods to be used for construction of such a minigene-bearing expression plasmid. Other expression vectors that may be used for minigene compositions are available and known to those of skill in the art.

The minigene DNA plasmid contains a consensus Kozak sequence and a consensus murine kappa Ig-light chain signal sequence followed by CTL and/or HTL epitopes selected in accordance with principles disclosed herein. The sequence encodes an open reading frame fused to the Myc and His antibody epitope tag coded for by the pcDNA 3.1 Myc-His vector.

Overlapping oligonucleotides, for example eight oligonucleotides, averaging approximately 70 nucleotides in length with 15 nucleotide overlaps, are synthesized and HPLC-purified. The oligonucleotides encode the selected peptide epitopes as well as appropriate linker nucleotides, Kozak sequence, and signal sequence. The final multiepitope minigene is assembled by extending the overlapping oligonucleotides in three sets of reactions using PCR. A Perkin/Elmer 9600 PCR machine is used and a total of 30 cycles are performed using the following conditions: 95°C for 15 sec, annealing temperature (5° below the lowest calculated Tm of each primer pair) for 30 sec, and 72°C for 1 min.

For the first PCR reaction, 5 μg of each of two oligonucleotides, *i.e.*, an amplification primer pair, are annealed and extended: Oligonucleotides 1+2, 3+4, 5+6, and 7+8 are combined in 100 μl reactions containing *Pfu* polymerase buffer (1x= 10 mM KCL, 10 mM $(NH_4)_2SO_4$, 20 mM Tris-chloride, pH 8.75, 2 mM $MgSO_4$, 0.1% Triton X-100, 100 μg/ml BSA), 0.25 mM each dNTP, and 2.5 U of *Pfu* polymerase. The full-length dimer products are gel-purified, and two reactions containing the product of 1+2 and 3+4, and the product of 5+6 and 7+8 are mixed, annealed, and extended for 10 cycles. Half of the two reactions are then mixed, and 5 cycles of annealing and extension carried out before flanking primers are added to amplify the full length product for 25 additional cycles. The full-length product is gel-purified and cloned into pCR-blunt (Invitrogen) and individual clones are screened by sequencing.

Example 12. The plasmid construct and the degree to which it induces immunogenicity.

The degree to which the plasmid construct prepared using the methodology outlined in Example 11 is able to induce immunogenicity is evaluated through *in vivo*

injections into mice and subsequent *in vitro* assessment of CTL and HTL activity, which are analysed using cytotoxicity and proliferation assays, respectively, as detailed *e.g.*, in U.S.S.N. 09/311,784 filed 5/13/99 and Alexander *et al.*, *Immunity* 1:751-761, 1994. For example, to assess the capacity of a pMin minigene construct that contains HLA-A2 supermotif epitopes to induce CTLs *in vivo*, HLA-A2.1/K^b transgenic mice are immunized intramuscularly with 100 μg of naked cDNA. As a means of comparing the level of CTLs induced by cDNA immunization, a control group of animals is also immunized with an actual peptide composition that comprises multiple epitopes synthesized as a single polypeptide as they would be encoded by the minigene.

Splenocytes from immunized animals are stimulated twice with each of the respective compositions (peptide epitopes encoded in the minigene or the polyepitopic peptide), then assayed for peptide-specific cytotoxic activity in a ^{51}Cr release assay. The results indicate the magnitude of the CTL response directed against the A3-restricted epitope, thus indicating the *in vivo* immunogenicity of the minigene vaccine and polyepitopic vaccine. It is, therefore, found that the minigene elicits immune responses directed toward the HLA-A2 supermotif peptide epitopes as does the polyepitopic peptide vaccine. A similar analysis is also performed using other HLA-A3 and HLA-B7 transgenic mouse models to assess CTL induction by HLA-A3 and HLA-B7 motif or supermotif epitopes.

To assess the capacity of a class II epitope encoding minigene to induce HTLs *in vivo*, I-A^b restricted mice, for example, are immunized intramuscularly with 100 μg of plasmid DNA. As a means of comparing the level of HTLs induced by DNA immunization, a group of control animals is also immunized with an actual peptide composition emulsified in complete Freund's adjuvant.

CD4+ T cells, *i.e.* HTLs, are purified from splenocytes of immunized animals and stimulated with each of the respective compositions (peptides encoded in the minigene). The HTL response is measured using a ^{3}H-thymidine incorporation proliferation assay, (*see*, *e.g.*, Alexander *et al.* *Immunity* 1:751-761, 1994). the results indicate the magnitude of the HTL response , thus demonstrating the *in vivo* immunogenicity of the minigene.

Alternatively, plasmid constructs can be evaluated *in vitro* by testing for epitope presentation by APC following transduction or transfection of the APC with an epitope-expressing nucleic acid construct. Such a study determines "antigenicity" and allows the use of human APC. The assay determines the ability of the epitope to be presented by the

APC in a context that is recognized by a T cell by quantifying the density of epitope-HLA class I complexes on the cell surface. Quantitation can be performed by directly measuring the amount of peptide eluted from the APC (*see, e.g.*, Sijts *et al.*, *J. Immunol.* 156:683-692, 1996; Demotz *et al.*, *Nature* 342:682-684, 1989); or the number of peptide-HLA class I complexes can be estimated by measuring the amount of lysis or lymphokine release induced by infected or transfected target cells, and then determining the concentration of peptide necessary to obtained equivalent levels of lysis or lymphokine release (*see, e.g.*, Kageyama *et al.*, *J. Immunol.* 154:567-576, 1995).

Example 13: Peptide Composition for Prophylactic Uses

Vaccine compositions of the present invention are used to prevent HCV infection in persons who are at risk for such infection. For example, a polyepitopic peptide epitope composition (or a nucleic acid comprising the same) containing multiple CTL and HTL epitopes such as those selected in Examples 9 and/or 10, which are also selected to target greater than 80% of the population, is administered to individuals at risk for HCV infection. The composition is provided as a single lipidated polypeptide that encompasses multiple epitopes. The vaccine is administered in an aqueous carrier comprised of Freunds Incomplete Adjuvant. The dose of peptide for the initial immunization is from about 1 to about 50,000 μg, generally 100-5,000 μg, for a 70 kg patient. The initial administration of vaccine is followed by booster dosages at 4 weeks followed by evaluation of the magnitude of the immune response in the patient, by techniques that determine the presence of epitope-specific CTL populations in a PBMC sample. Additional booster doses are administered as required. The composition is found to be both safe and efficacious as a prophylaxis against HCV infection.

Alternatively, the polyepitopic peptide composition can be administered as a nucleic acid in accordance with methodologies known in the art and disclosed herein.

Example 14: Polyepitopic Vaccine Compositions Derived from Native HCV Sequences

A native HCV polyprotein sequence is screened, preferably using computer algorithms defined for each class I and/or class II supermotif or motif, to identify "relatively short" regions of the polyprotein that comprise multiple epitopes and is preferably less in length than an entire native antigen. This relatively short sequence that contains multiple distinct, even overlapping, epitopes is selected and used to generate a minigene construct. The construct is engineered to express the peptide, which

corresponds to the native protein sequence. The "relatively short" peptide is generally less than 250 amino acids in length, often less than 100 amino acids in length, preferably less than 75 amino acids in length, and more preferably less than 50 amino acids in length. The protein sequence of the vaccine composition is selected because it has maximal number of epitopes contained within the sequence, *i.e.*, it has a high concentration of epitopes. As noted herein, epitope motifs may be nested or overlapping (*i.e.*, frame shifted relative to one another). For example, with frame shifted overlapping epitopes, two 9-mer epitopes and one 10-mer epitope can be present in a 10 amino acid peptide. Such a vaccine composition is administered for therapeutic or prophylactic purposes.

The vaccine composition will preferably include, for example, three CTL epitopes and at least one HTL epitope from an HCV antigen. This polyepitopic native sequence is administered either as a peptide or as a nucleic acid sequence which encodes the peptide. Alternatively, an analog can be made of this native sequence, whereby one or more of the epitopes comprise substitutions that alter the cross-reactivity and/or binding affinity properties of the polyepitopic peptide.

The embodiment of this example provides for the possibility that an as yet undiscovered aspect of immune system processing will apply to the native nested sequence and thereby facilitate the production of therapeutic or prophylactic immune response-inducing vaccine compositions. Additionally such an embodiment provides for the possibility of motif-bearing epitopes for an HLA makeup that is presently unknown. Furthermore, this embodiment (absent analogs) directs the immune response to multiple peptide sequences that are actually present in native HCV antigens thus avoiding the need to evaluate any junctional epitopes. Lastly, the embodiment provides an economy of scale when producing nucleic acid vaccine compositions.

Related to this embodiment, computer programs can be derived in accordance with principles in the art, which identify in a target sequence, the greatest number of epitopes per sequence length.

<u>Example 15. Polyepitopic Vaccine Compositions Directed To Multiple Diseases</u>

The HCV peptide epitopes of the present invention are used in conjunction with peptide epitopes from target antigens related to one or more other diseases, to create a vaccine composition that is useful for the prevention or treatment of HCV as well as the

one or more other disease(s). Examples of the other diseases include, but are not limited to, HIV, and HBV.

For example, a polyepitopic peptide composition comprising multiple CTL and HTL epitopes that target greater than 98% of the population may be created for administration to individuals at risk for both HCV and HIV infection. The composition can be provided as a single polypeptide that incorporates the multiple epitopes from the various disease-associated sources, or can be administered as a composition comprising one or more discrete epitopes.

Example 16. Use of peptides to evaluate an immune response

Peptides of the invention may be used to analyze an immune response for the presence of specific CTL or HTL populations directed to a prostate cancer-associated antigen. Such an analysis may be performed using multimeric complexes as described, *e.g.*, by Ogg *et al., Science* 279:2103-2106, 1998 and Greten *et al.*, *Proc. Natl. Acad. Sci. USA* 95:7568-7573, 1998. In the following example, peptides in accordance with the invention are used as a reagent for diagnostic or prognostic purposes, not as an immunogen.

In this example, highly sensitive human leukocyte antigen tetrameric complexes ("tetramers") are used for a cross-sectional analysis of, for example, HCV HLA-A*0201-specific CTL frequencies from HLA A*0201-positive individuals at different stages of disease or following immunization using an HCV peptide containing an A*0201 motif. Tetrameric complexes are synthesized as described (Musey *et al., N. Engl. J. Med.* 337:1267, 1997). Briefly, purified HLA heavy chain (A*0201 in this example) and β2-microglobulin are synthesized by means of a prokaryotic expression system. The heavy chain is modified by deletion of the transmembrane-cytosolic tail and COOH-terminal addition of a sequence containing a BirA enzymatic biotinylation site. The heavy chain, β2-microglobulin, and peptide are refolded by dilution. The 45-kD refolded product is isolated by fast protein liquid chromatography and then biotinylated by BirA in the presence of biotin (Sigma, St. Louis, Missouri), adenosine 5'triphosphate and magnesium. Streptavidin-phycoerythrin conjugate is added in a 1:4 molar ratio, and the tetrameric product is concentrated to 1 mg/ml. The resulting product is referred to as tetramer-phycoerythrin.

For the analysis of patient blood samples, approximately one million PBMCs are centrifuged at 300g for 5 minutes and resuspended in 50 µl of cold phosphate-buffered saline. Tri-color analysis is performed with the tetramer-phycoerythrin, along with anti-CD8-Tricolor, and anti-CD38. The PBMCs are incubated with tetramer and antibodies on ice for 30 to 60 min and then washed twice before formaldehyde fixation. Gates are applied to contain >99.98% of control samples. Controls for the tetramers include both A*0201-negative individuals and A*0201-positive uninfected donors. The percentage of cells stained with the tetramer is then determined by flow cytometry. The results indicate the number of cells in the PBMC sample that contain epitope-restricted CTLs, thereby readily indicating the extent of immune response to the HCV epitope, and thus the stage of HCV infection or exposure to a vaccine that elicits a protective or therapeutic response.

Example 17: Use of Peptide Epitopes to Evaluate Recall Responses

The peptide epitopes of the invention are used as reagents to evaluate T cell responses, such as acute or recall responses, in patients. Such an analysis may be performed on patients who have recovered from infection, who are chronically infected with HCV, or who have been vaccinated with an HCV vaccine.

For example, the class I restricted CTL response of persons who have been vaccinated may be analyzed. The vaccine may be any HCV vaccine. PBMC are collected from vaccinated individuals and HLA typed. Appropriate peptide epitopes of the invention that are preferably highly conserved and, optimally, bear supermotifs to provide cross-reactivity with multiple HLA supertype family members, are then used for analysis of samples derived from individuals who bear that HLA type.

PBMC from vaccinated individuals are separated on Ficoll-Histopaque density gradients (Sigma Chemical Co., St. Louis, MO), washed three times in HBSS (GIBCO Laboratories), resuspended in RPMI-1640 (GIBCO Laboratories) supplemented with L-glutamine (2mM), penicillin (50U/ml), streptomycin (50 µg/ml), and Hepes (10mM) containing 10% heat-inactivated human AB serum (complete RPMI) and plated using microculture formats. A synthetic peptide comprising an epitope of the invention is added at 10 µg/ml to each well and HBV core 128-140 epitope is added at 1 µg/ml to each well as a source of T cell help during the first week of stimulation.

In the microculture format, 4 x 10^5 PBMC are stimulated with peptide in 8 replicate cultures in 96-well round bottom plate in 100 µl/well of complete RPMI. On

days 3 and 10, 100 ml of complete RPMI and 20 U/ml final concentration of rIL-2 are added to each well. On day 7 the cultures are transferred into a 96-well flat-bottom plate and restimulated with peptide, rIL-2 and 10^5 irradiated (3,000 rad) autologous feeder cells. The cultures are tested for cytotoxic activity on day 14. A positive CTL response requires two or more of the eight replicate cultures to display greater than 10% specific ^{51}Cr release, based on comparison with uninfected control subjects as previously described (Rehermann, *et al., Nature Med.* 2:1104,1108, 1996; Rehermann *et al., J. Clin. Invest.* 97:1655-1665, 1996; and Rehermann *et al. J. Clin. Invest.* 98:1432-1440, 1996).

Target cell lines are autologous and allogeneic EBV-transformed B-LCL that are either purchased from the American Society for Histocompatibility and Immunogenetics (ASHI, Boston, MA) or established from the pool of patients as described (Guilhot, *et al. J. Virol.* 66:2670-2678, 1992).

Cytotoxicity assays are performed in the following manner. Target cells consist of either allogeneic HLA-matched or autologous EBV-transformed B lymphoblastoid cell line that are incubated overnight with the synthetic peptide epitope of the invention at 10 μM, and labeled with 100 μCi of ^{51}Cr (Amersham Corp., Arlington Heights, IL) for 1 hour after which they are washed four times with HBSS.

Cytolytic activity is determined in a standard 4-h, split well ^{51}Cr release assay using U-bottomed 96 well plates containing 3,000 targets/well. Stimulated PBMC are tested at effector/target (E/T) ratios of 20-50:1 on day 14. Percent cytotoxicity is determined from the formula: 100 x [(experimental release-spontaneous release)/maximum release-spontaneous release)]. Maximum release is determined by lysis of targets by detergent (2% Triton X-100; Sigma Chemical Co., St. Louis, MO). Spontaneous release is <25% of maximum release for all experiments.

The results of such an analysis indicate the extent to which HLA-restricted CTL populations have been stimulated by previous exposure to HCV or an HCV vaccine.

The class II restricted HTL responses may also be analyzed. Purified PBMC are cultured in a 96-well flat bottom plate at a density of 1.5×10^5 cells/well and are stimulated with 10 μg/ml synthetic peptide, whole antigen, or PHA. Cells are routinely plated in replicates of 4-6 wells for each condition. After seven days of culture, the medium is removed and replaced with fresh medium containing 10U/ml IL-2. Two days later, 1 μCi ^{3}H-thymidine is added to each well and incubation is continued for an additional 18 hours. Cellular DNA is then harvested on glass fiber mats and analyzed for ^{3}H-thymidine

incorporation. Antigen-specific T cell proliferation is calculated as the ratio of ^{3}H-thymidine incorporation in the presence of antigen divided by the ^{3}H-thymidine incorporation in the absence of antigen.

Example 18: Induction Of Specific CTL Response In Humans

A human clinical trial for an immunogenic composition comprising CTL and HTL epitopes of the invention is set up as an IND Phase I, dose escalation study and carried out as a randomized, double-blind, placebo-controlled trial. Such a trial is designed, for example, as follows:

A total of about 27 subjects are enrolled and divided into 3 groups:

Group I: 3 subjects are injected with placebo and 6 subjects are injected with 5 μg of peptide composition;

Group II: 3 subjects are injected with placebo and 6 subjects are injected with 50 μg peptide composition;

Group III: 3 subjects are injected with placebo and 6 subjects are injected with 500 μg of peptide composition.

After 4 weeks following the first injection, all subjects receive a booster inoculation at the same dosage.

The endpoints measured in this study relate to the safety and tolerability of the peptide composition as well as its immunogenicity. Cellular immune responses to the peptide composition are an index of the intrinsic activity of this the peptide composition, and can therefore be viewed as a measure of biological efficacy. The following summarize the clinical and laboratory data that relate to safety and efficacy endpoints.

Safety: The incidence of adverse events is monitored in the placebo and drug treatment group and assessed in terms of degree and reversibility.

Evaluation of Vaccine Efficacy: For evaluation of vaccine efficacy, subjects are bled before and after injection. Peripheral blood mononuclear cells are isolated from fresh heparinized blood by Ficoll-Hypaque density gradient centrifugation, aliquoted in freezing media and stored frozen. Samples are assayed for CTL and HTL activity.

The vaccine is found to be both safe and efficacious.

Example 19: Phase II Trials In Patients Infected With HCV

Phase II trials are performed to study the effect of administering the CTL-HTL peptide compositions to patients having chronic HCV infection. The main objectives of

the trials are to determine an effective dose and regimen for inducing CTLs in chronically infected HCV patients, to establish the safety of inducing a CTL and HTL response in these patients, and to see to what extent activation of CTLs improves the clinical picture of chronically infected CTL patients, as manifested by a transient flare in alanine aminotransferase (ALT), normalization of ALT, and reduction in HCV DNA. Such a study is designed, for example, as follows:

The studies are performed in multiple centers. The trial design is an open-label, uncontrolled, dose escalation protocol wherein the peptide composition is administered as a single dose followed six weeks later by a single booster shot of the same dose. The dosages are 50, 500 and 5,000 micrograms per injection. Drug-associated adverse effects (severity and reversibility) are recorded.

There are three patient groupings. The first group is injected with 50 micrograms of the peptide composition and the second and third groups with 500 and 5,000 micrograms of peptide composition, respectively. The patients within each group range in age from 21-65, include both males and females, and represent diverse ethnic backgrounds. All of them are infected with HCV for over five years and are HIV, HBV and delta hepatitis virus (HDV) negative, but have positive levels of HCV antigen.

The magnitude and incidence of ALT flares and the levels of HCV DNA in the blood are monitored to assess the effects of administering the peptide compositions. The levels of HCV DNA in the blood are an indirect indication of the progress of treatment. The vaccine composition is found to be both safe and efficacious in the treatment of chronic HCV infection.

Example 20. Induction of CTL Responses Using a Prime Boost Protocol

A prime boost protocol can also be used for the administration of the vaccine to humans. Such a vaccine regimen may include an initial administration of, for example, naked DNA followed by a boost using recombinant virus encoding the vaccine, or recombinant protein/polypeptide or a peptide mixture administered in an adjuvant.

For example, the initial immunization may be performed using an expression vector, such as that constructed in Example 11, in the form of naked nucleic acid administered IM (or SC or ID) in the amounts of 0.5-5 mg at multiple sites. The nucleic acid (0.1 to 1000 μg) can also be administered using a gene gun. Following an incubation period of 3-4 weeks, a booster dose is administered. The booster can, *e.g.*, be recombinant fowlpox virus administered at a dose of $5\text{-}10^7$ to 5×10^9 pfu. An alternative

recombinant virus, such as an MVA, canarypox, adenovirus, or adeno-associated virus, can also be used for the booster, or the polyepitopic protein or a mixture of the peptides can be administered. For evaluation of vaccine efficacy, patient blood samples will be obtained before immunization as well as at intervals following administration of the initial vaccine and booster doses of the vaccine. Peripheral blood mononuclear cells are isolated from fresh heparinized blood by Ficoll-Hypaque density gradient centrifugation, aliquoted in freezing media and stored frozen. Samples are assayed for CTL and HTL activity.

Analysis of the results will indicate that a magnitude of response sufficient to achieve protective immunity or to treat HCV infection infection is generated.

Example 21. Administration of Vaccine Compositions Using Dendritic Cells

Vaccines comprising peptide epitopes of the invention may be administered using dendritic cells. In this example, the peptide-pulsed dendritic cells can be administered to a patient to stimulate a CTL response *in vivo*. In this method dendritic cells are isolated, expanded, and pulsed with a vaccine comprising peptide CTL and HTL epitopes of the invention. The dendritic cells are infused back into the patient to elicit CTL and HTL responses *in vivo*. The induced CTL and HTL then destroy (CTL) or facilitate destruction (HTL) of the specific target HCV-infected cells that bear the proteins from which the epitopes in the vaccine are derived.

Alternatively, *Ex vivo* CTL or HTL responses to a particular tumor-associated antigen can be induced by incubating in tissue culture the patient's, or genetically compatible, CTL or HTL precursor cells together with a source of antigen-presenting cells, such as dendritic cells, and the appropriate immunogenic peptides. After an appropriate incubation time (typically about 7-28 days), in which the precursor cells are activated and expanded into effector cells, the cells are infused back into the patient, where they will destroy (CTL) or facilitate destruction (HTL) of their specific target cells, *i.e.*, tumor cells.

Example 22: Alternative Method of Identifying Motif-Bearing Peptides

Another way of identifying motif-bearing peptides is to elute them from cells bearing defined MHC molecules. For example, EBV transformed B cell lines used for tissue typing, have been extensively characterized to determine which HLA molecules they express. In certain cases these cells express only a single type of HLA molecule.

These cells can then be infected with a pathogenic organism, *e.g.*, HCV, or transfected with nucleic acids that express the antigen of interest. Thereafter, peptides produced by endogenous antigen processing of peptides produced consequent to infection (or as a result of transfection) will bind be displayed on the cell surface. The peptides are then eluted from the HLA molecules by exposure to mild acid conditions and their amino acid sequence determined, e.g., by mass spectral analysis (*e.g.*, Kubo *et al.*, *J. Immunol.* 152:3913, 1994). Because, as disclosed herein, the majority of peptides that bind a particular HLA molecule are motif-bearing, this is an alternative modality for obtaining the motif-bearing peptides correlated with the particular HLA molecule expressed on the cell.

Alternatively, cell lines that do not express any endogenous HLA molecules can be transfected with an expression construct encoding a single HLA allele. These cells may then be used as described, *i.e.*, they may be infected with a pathogenic organism or transfected with nucleic acid encoding an antigen of interest to isolate peptides corresponding to the pathogen or antigen of interest that have been presented on the cell surface. Peptides obtained from such an analysis will bear motif(s) that correspond to binding to the single HLA allele that is expressed in the cell.

As appreciated by one in the art, one can perform a similar analysis on a cell bearing more than one HLA allele and subsequently determine peptides specific for each HLA allele expressed. Moreover, one of skill would also recognize that means other than infection or transfection, such as loading with a protein antigen, can be used to provide a source of antigen to the cell.

The above examples are provided to illustrate the invention but not to limit its scope. For example, the human terminology for the Major Histocompatibility Complex, namely HLA, is used throughout this document. It is to be appreciated that these principles can be extended to other species as well. Thus, other variants of the invention will be readily apparent to one of ordinary skill in the art and are encompassed by the appended claims. All publications, patents, and patent application cited herein are hereby incorporated by reference for all purposes.

TABLE I

SUPERMOTIFS	POSITION 2 (Primary Anchor)	POSITION 3 (Primary Anchor)	POSITION C Terminus (Primary Anchor)
A1	**T, I,** *L, V, M, S*		**F, W, Y**
A2	**L, I, V, M,** *A, T, Q*		**I, V,** *M, A, T, L*
A3	**V, S, M, A,** *T, L, I*		**R,K**
A24	**Y, F,** *W, I, V, L, M, T*		**F, I,** *Y, W, L, M*
B7	**P**		**V, I, L, F,** *M, W, Y, A*
B27	**R, H, K**		**F, Y, L,** *W, M, I, V, A*
B44	**E,** *D*		**F, W, L, I, M, V, A**
B58	**A, T, S**		**F, W, Y,** *L, I, V, M, A*
B62	**Q, L,** *I, V, M, P*		**F, W, Y,** *M, I, V, L, A*
MOTIFS			
A1	**T, S, M**		**Y**
A1		**D, E,** *A, S*	**Y**
A2.1	**L, M,** *V, Q, I, A, T*		**V,** *L, I, M, A, T*
A3	**L, M, V, I, S, A, T, F,** *C, G, D*		**K, Y, R,** *H, F, A*
A11	**V, T, M, L, I, S, A, G, N,** *C, D, F*		**K,** *R, Y, H*
A24	**Y, F, W,** *M*		**F, L, I, W**
A*3101	**M, V, T,** *A ,L, I, S*		**R,** *K*
A*3301	**M, V, A, L, F,** *I, S, T*		**R, K**
A*6801	**A, V, T,** *M, S, L, I*		**R, K**
B*0702	**P**		**L, M, F,** *W, Y, A, I, V*
B*3501	**P**		**L, M, F, W, Y,** *I, V, A*
B51	**P**		**L, I, V, F,** *W, Y, A, M*
B*5301	**P**		**I, M, F, W, Y,** *A, L, V*
B*5401	**P**		**A, T, I, V,** *L, M, F, W, Y*

Bolded residues are preferred, italicized residues are less preferred: A peptide is considered motif-bearing if it has primary anchors at each primary anchor position for a motif or supermotif as specified in the above table.

TABLE II

		POSITION								
		1	2	3	4	5	6	7	8	C-terminus
<u>SUPERMOTIFS</u>										
A1			<u>1° Anchor</u> T,I,*L,V,M,S*							<u>1° Anchor</u> F,W,Y
A2			<u>1° Anchor</u> L,I,V,M,*A,T,Q*							<u>1° Anchor</u> L,I,V,M,A,T
A3	preferred		<u>1° Anchor</u> V,S,M,A,*T,L,I*	Y,F,W (4/5)			Y,F,W (3/5)	Y,F,W (4/5)	P (4/5)	<u>1°Anchor</u> R,K
	deleterious	D,E (3/5); P (5/5)		D,E (4/5)						
A24			<u>1° Anchor</u> Y,F,*W,I,V,L,M,T*							<u>1° Anchor</u> F,I,*Y,W,L,M*
B7	preferred	F,W,Y (5/5) L,I,V,M (3/5)	<u>1°Anchor</u> P	F,W,Y (4/5)					F,W,Y (3/5)	<u>1°Anchor</u> V,I,L,F,*M,W,Y,A*
	deleterious	D,E (3/5); P(5/5); G(4/5); A(3/5); Q,N (3/5)				D,E (3/5)	G (4/5)	Q,N (4/5)	D,E (4/5)	
B27			<u>1° Anchor</u> R,H,K							<u>1° Anchor</u> F,Y,L,*W,M,V,A*
B44			<u>1° Anchor</u> E,*D*							<u>1° Anchor</u> F,W,Y,L,I,M,V,A
B58			<u>1° Anchor</u> A,T,S							<u>1° Anchor</u> F,W,Y,*L,I,V,M,A*
B62			<u>1° Anchor</u> Q,L,*I,V,M,P*							<u>1° Anchor</u> F,W,Y,*M,I,V,L,A*

		POSITION								
		1	2	3	4	5	6	7	8	C-terminus
MOTIFS										
A1 9-mer	preferred	G,F,Y,W	1°Anchor S,T,M	D,E,A	Y,F,W		P	D,E,Q,N	Y,F,W	1°Anchor Y
	deleterious	D,E		R,H,K,L,I,V M,P	A	G	A			
A1 9-mer	preferred	G,R,H,K	A,S,T,C,L,I V,M,	1°Anchor D,E,*A*,*S*	G,S,T,C		A,S,T,C	L,I,V,M	D,E	1°Anchor Y
	deleterious	A	R,H,K,D,E, P,Y,F,W		D,E	P,Q,N	R,H,K	P,G	G,P	

		POSITION									
		1	2	3	4	5	6	7	8	9 or C-terminus	C-terminus
A1 10-mer	peferred	Y,F,W	<u>1°Anchor</u> S,T,M	D,E,A,Q,N	A	Y,F,W,Q,N		P,A,S,T,C	G,D,E	P	<u>1°Anchor</u> Y
	deleterious	G,P		R,H,K,G,L,I V,M	D,E	R,H,K	Q,N,A	R,H,K,Y,F, W	R,H,K	A	
A1 10-mer	preferred	Y,F,W	S,T,C,L,I,V M	<u>1°Anchor</u> D,E,*A*,*S*	A	Y,F,W		P,G	G	Y,F,W	<u>1°Anchor</u> Y
	deleterious	R,H,K	R,H,K,D,E, P,Y,F,W			P	G		P,R,H,K	Q,N	
A2.1 9-mer	preferred	Y,F,W	<u>1°Anchor</u> L,M,*I*,*V*,*Q*, *A*,*T*	Y,F,W	S,T,C	Y,F,W		A	P	<u>1°Anchor</u> V,*L*,*I*,*M*,*A*,*T*	
	deleterious	D,E,P		D,E,R,K,H			R,K,H	D,E,R,K,H			
A2.1 10-mer	preferred	A,Y,F,W	<u>1°Anchor</u> L,M,*I*,*V*,*Q*, *A*,*T*	L,V,I,M	G		G		F,Y,W, L,V,I,M		<u>1°Anchor</u> V,*L*,*I*,*M*,*A*,*T*
	deleterious	D,E,P		D,E	R,K,H,A	P		R,K,H	D,E,R, K,H	R,K,H	

		POSITION									
		1	2	3	4	5	6	7	8	9 or C-terminus	C-terminus
A3	preferred	R,H,K	<u>1°Anchor</u> L,M,V,I,S, A,T,F,*C*,*G* *D*	Y,F,W	P,R,H,K,Y, F,W	A	Y,F,W		P	<u>1°Anchor</u> K,Y,R,*H*,*F*,*A*	
	deleterious	D,E,P		D,E							
A11	preferred	A	<u>1°Anchor</u> V,T,L,M,I, S,A,G,N,*C*, *D*,*F*	Y,F,W	Y,FW	A	Y,F,W	Y,FW	P	<u>1°Anchor</u> K,,*RY*,*H*	
	deleterious	D,E,P						A	G		
A24 9-mer	preferred	Y,F,W,R,H,K	<u>1°Anchor</u> Y,F,W,*M*		S,T,C			Y,F,W	Y,F,W	<u>1°Anchor</u> F,L,I,W	
	deleterious	D,E,G		D,E	G	Q,N,P	D,E,R,H,K	G	A,Q,N		
A24 10-mer	preferred		<u>1°Anchor</u> Y,F,W,*M*		P	Y,F,W,P		P			<u>1°Anchor</u> F,L,I,W
	deleterious			G,D,E	Q,N	R,H,K	D,E	A	Q,N	D,E,A	
A3101	preferred	R,H,K	<u>1°Anchor</u> M,V,T,*A*,*L*, *I*,*S*	Y,F,W	P		Y,F,W	Y,F,W	A,P	<u>1°Anchor</u> R,*K*	
	deleterious	D,E,P		D,E		A,D,E	D,E	D,E	D,E		

		POSITION									
		1	2	3	4	5	6	7	8	9 or C-terminus	C-terminus
A3301	preferred		<u>1°Anchor</u> M,V,A,L,F, *I,S,T*	Y,F,W				A,Y,F,W		<u>1°Anchor</u> R,K	
	deleterious	G,P		D,E							
A6801	preferred	Y,F,W,S,T,C	<u>1°Anchor</u> A,V,T,*M,S, L,I*			Y,F,W,L,I, V,M		Y,F,W	P	<u>1°Anchor</u> R,K	
	deleterious	G,P		D,E,G		R,H,K			A		
B0702	preferred	R,H,K,F,W,Y	<u>1°Anchor</u> P	R,H,K		R,H,K	R,H,K	R,H,K	P,A	<u>1°Anchor</u> L,M,F,*W,Y,A, I,V*	
	deleterious	D,E,Q,N,P		D,E,P	D,E	D,E	G,D,E	Q,N	D,E		
B3501	preferred	F,W,Y,L,I,V,M	<u>1°Anchor</u> P	F,W,Y				F,W,Y		<u>1°Anchor</u> L,M,F,W,Y,*I, V,A*	
	deleterious	A,G,P				G	G				

		POSITION									
		1	2	3	4	5	6	7	8	9 or C-terminus	C-terminus
B51	preferred	L,I,V,M,F,W,Y	<u>1°Anchor</u> P	F,W,Y	S,T,C	F,W,Y		G	F,W,Y	<u>1°Anchor</u> L,I,V,F,*W,* *Y,A,M*	
	deleterious	A,G,P,D,E,R,H,K, S,T,C				D,E	G	D,E,Q,N	G,D,E		
B5301	preferred	L,I,V,M,F,W,Y	<u>1°Anchor</u> P	F,W,Y	S,T,C	F,W,Y		L,I,V,M,F, W,Y	F,W,Y	<u>1°Anchor</u> I,M,F,W,Y, *A,L,V*	
	deleterious	A,G,P,Q,N					G	R,H,K,Q,N	D,E		
B5401	preferred	F,W,Y	<u>1°Anchor</u> P	F,W,Y,L,I,V M		L,I,V,M		A,L,I,V,M	F,W,Y,A,P	<u>1°Anchor</u> A,T,I,V,*L,* *M,F,W,Y*	
	deleterious	G,P,Q,N,D,E		G,D,E,S,T,C		R,H,K,D,E	D,E	Q,N,D,G,E	D,E		

Italicized residues indicate less preferred or "tolerated" residues.
The information in Table II is specific for 9-mers unless otherwise specified.

Table III

POSITION

MOTIFS		1° anchor 1	2	3	4	5	1° anchor 6	7	8	9
DR4	preferred	F, M, Y, *L, I, V, W*	M	T		I	V, S, T, *C, P, A, L, I, M*	M, H,		M, H
	deleterious				W,			R,		W, D, E
DR1	preferred	M, F, *L, I, V, W, Y*			P, A, M, Q		V, M, A, T, *S, P, L, I, C*	M,		A, V, M
	deleterious		C	C, H	F, D	C, W, D		G, D, E,	D	
DR7	preferred	M, F, *L, I, V, W, Y*	M	W	A		I, V, M, S, A, *C, T, P, L*	M		I, V
	deleterious		C,		G,			G, R, D	N	G
DR Supermotif		M, F, *L, I, V, W, Y*					V, M, S, T, A, *C, P, L, I*			
DR3 MOTIFS		1° anchor 1	2	3	1° anchor 4	5	1° anchor 6			
motif a preferred		L, I, V, M, F, Y			D					
motif b preferred		L, I, V, M, F, A, Y			D, N, Q, E, S, T		K, R, H			

Italicized residues indicate less preferred or "tolerated" residues.

Table IV: HLA Class I Standard Peptide Binding Affinity.

ALLELE	STANDARD PEPTIDE	SEQUENCE (SEQ ID NO:)	STANDARD BINDING AFFINITY (nM)
A*0101	944.02	YLEPAIAKY	25
A*0201	941.01	FLPSDYFPSV	5.0
A*0202	941.01	FLPSDYFPSV	4.3
A*0203	941.01	FLPSDYFPSV	10
A*0205	941.01	FLPSDYFPSV	4.3
A*0206	941.01	FLPSDYFPSV	3.7
A*0207	941.01	FLPSDYFPSV	23
A*6802	1072.34	YVIKVSARV	8.0
A*0301	941.12	KVFPYALINK	11
A*1101	940.06	AVDLYHFLK	6.0
A*3101	941.12	KVFPYALINK	18
A*3301	1083.02	STLPETYVVRR	29
A*6801	941.12	KVFPYALINK	8.0
A*2402	979.02	AYIDNYNKF	12
B*0702	1075.23	APRTLVYLL	5.5
B*3501	1021.05	FPFKYAAAF	7.2
B51	1021.05	FPFKYAAAF	5.5
B*5301	1021.05	FPFKYAAAF	9.3
B*5401	1021.05	FPFKYAAAF	10

Table V. HLA Class II Standard Peptide Binding Affinity.

Allele	Nomenclature	Standard Peptide	Sequence (SEQ ID NO:)	Binding Affinity (nM)
DRB1*0101	DR1	515.01	PKYVKQNTLKLAT	5.0
DRB1*0301	DR3	829.02	YKTIAFDEEARR	300
DRB1*0401	DR4w4	515.01	PKYVKQNTLKLAT	45
DRB1*0404	DR4w14	717.01	YARFQSQTTLKQKT	50
DRB1*0405	DR4w15	717.01	YARFQSQTTLKQKT	38
DRB1*0701	DR7	553.01	QYIKANSKFIGITE	25
DRB1*0802	DR8w2	553.01	QYIKANSKFIGITE	49
DRB1*0803	DR8w3	553.01	QYIKANSKFIGITE	1600
DRB1*0901	DR9	553.01	QYIKANSKFIGITE	75
DRB1*1101	DR5w11	553.01	QYIKANSKFIGITE	20
DRB1*1201	DR5w12	1200.05	EALIHQLKINPYVLS	298
DRB1*1302	DR6w19	650.22	QYIKANAKFIGITE	3.5
DRB1*1501	DR2w2β1	507.02	GRTQDENPVVHFFKNIV TPRTPPP	9.1
DRB3*0101	DR52a	511	NGQIGNDPNRDIL	470
DRB4*0101	DRw53	717.01	YARFQSQTTLKQKT	58
DRB5*0101	DR2w2β2	553.01	QYIKANSKFIGITE	20

Table VI

HLA-supertype	Allelle-specific HLA-supertype members	
	Verified[a]	Predicted[b]
A1	A*0101, A*2501, A*2601, A*2602, A*3201	A*0102, A*2604, A*3601, A*4301, A*8001
A2	A*0201, A*0202, A*0203, A*0204, A*0205, A*0206, A*0207, A*0209, A*0214, A*6802; A*6901	A*0208, A*0210, A*0211, A*0212, A*0213
A3	A*0301, A*1101, A*3101, A*3301, A*6801	A*0302, A*1102, A*2603, A*3302, A*3303, A*3401, A*3402, A*6601, A*6602, A*7401
A24	A*2301, A*2402, A*3001	A*2403, A*2404, A*3002, A*3003
B7	B*0702, B*0703, B*0704, B*0705, B*1508, B*3501, B*3502, B*3503, B*3503, B*3504, B*3505, B*3506, B*3507, B*3508, B*5101, B*5102, B*5103, B*5104, B*5105, B*5301, B*5401, B*5501, B*5502, B*5601, B*5602, B*6701, B*7801	B*1511, B*4201, B*5901
B27	B*1401, B*1402, B*1509, B*2702, B*2703, B*2704, B*2705, B*2706, B*3801, B*3901, B*3902, B*7301	B*2701, B*2707, B*2708, B*3802, B*3903, B*3904, B*3905, B*4801, B*4802, B*1510, B*1518, B*1503
B44	B*1801, B*1802, B*3701, B*4402, B*4403, B*4404, B*4001, B*4002, B*4006	B*4101, B*4501, B*4701, B*4901, B*5001
B58	B*5701, B*5702, B*5801, B*5802, B*1516, B*1517	
B62	B*1501, B*1502, B*1513, B*5201	B*1301, B*1302, B*1504, B*1505, B*1506, B*1507, B*1515, B*1520, B*1521, B*1512, B*1514, B*1510

a. Verified alleles include alleles whose specificity has been determined by pool sequencing analysis, peptide binding assays, or by analysis of the sequences of CTL epitopes.

b. Predicted alleles are alleles whose specificity is predicted on the basis of B and F pocket structure to overlap with the supertype specificity.

Table VII

HCV A01 Super Motif with Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0101
ATGNLPGCSF	185	10	13	93	
ATLGFGAY	1265	8	14	100	
AVQWMNRLIAF	1917	11	14	100	
CTCGSSDLY	1128	9	11	79	0.3700
CTRGVAKAVDF	1190	11	11	79	
CTWMNSTGF	555	9	11	79	
CVTQTVDF	1462	8	12	86	
DLEVVTSTW	1857	9	12	86	
ETTMRSPVF	1207	9	12	86	
FSYDTRCF	2870	8	11	79	
FTEAMTRY	2792	8	14	100	
FTGLTHIDAHF	1567	11	13	93	
GLPVCQDHLEF	1552	11	12	86	
GLSAFSLHSY	2921	10	11	79	0.0029
GLTHIDAHF	1569	9	13	93	
GSSYGFQY	2841	8	11	79	
GTFPINAY	2063	8	11	79	
GVAGALVAF	1863	9	12	86	
GVAKAVDF	1193	8	11	79	
GVLAALAAY	1670	9	12	86	
GVRVCEKMALY	2619	11	14	100	
GVRVLEDGVNY	154	11	12	86	
HLIQNIVDVQY	696	11	11	79	
HMWNFISGIQY	1769	11	13	93	
HVGPGEGAVQW	1910	11	11	79	
IMAKNEVF	2591	8	12	86	
ITYSTYGKF	1296	9	12	86	
IVDVQYLY	701	8	12	86	
KSTKVPAAY	1241	9	12	86	0.0130
KVIDTLTCGF	121	10	12	86	
LIEANLLW	2235	8	12	86	
LINTNGSW	414	8	11	79	
LLAPITAY	1030	8	14	100	
LLFNILGGW	1812	9	12	86	
LLSPRGSRPSW	97	11	11	79	
LSAFSLHSY	2922	9	11	79	0.8100
LSPRGSRPSW	98	10	11	79	
LTCGFADLMGY	126	11	12	86	
LTHIDAHF	1570	8	13	93	
LVDILAGY	1853	8	11	79	
MILMTHFF	2878	8	12	86	
NIVDVQYLY	700	9	12	86	0.0980
NLPGCSFSIF	168	10	13	93	
NTCVTQTVDF	1400	10	12	86	
NINRRPQDVKF	14	11	11	79	

HCV A01 Super Motif with Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0101
NVDQDLVGW	1108	9	11	79	
PITYSTYGKF	1295	10	11	79	
PMGFSYDTRCF	2667	11	11	79	
PSVAATLGF	1261	9	14	100	
PTLHGPTPLLY	1621	11	11	79	
PVCQDHLEF	1554	9	12	86	
PVCQDHLEFW	1554	10	12	86	
QTVDFSLDPTF	1485	11	12	86	
RLHGLSAF	2918	8	12	86	
RLLAPITAY	1029	9	12	86	
RMAWDMMMNW	317	10	12	86	
RMILMTHF	2875	8	12	86	
RMILMTHFF	2875	9	12	86	
RVCEKMALY	2621	9	14	100	
RVLEDGVNY	156	9	12	86	
STKVPAAY	1242	8	12	86	
SVAATLGF	1262	8	14	100	
SVAATLGFGAY	1262	11	14	100	
TIMAKNEVF	2590	9	11	79	
TLHGPTPLLY	1622	10	11	79	0.0300
TLLFNILGGW	1811	10	12	86	
TTIMAKNEVF	2589	10	11	79	
TTMRSPVF	1208	8	12	86	
TVDFSLDPTF	1486	10	12	86	
VIDTLTCGF	122	9	12	86	
VLAALAAY	1871	8	12	86	
VLEDGVNY	157	8	12	86	
VLVDILAGY	1052	9	11	79	
VMGSSYGF	2639	8	11	79	
VMGSSYGFQY	2639	10	11	79	
WMNRLIAF	1920	8	14	100	
YSPGQRVEF	2848	9	11	79	
YTNVDQDLVGW	1106	11	11	79	
YVGDLCGSVF	276	10	12	86	
79		2			

Table VIII

HCV A02 Super Motif with Binding Information

Conservancy	Freq.	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
93	13	1904	AAILRRHV					
86	12	1673	AALAAYCL					
79	11	1250	AAQGYKVL					
79	11	1250	AAQGYKVLV					
79	11	1250	AAQGYKVLVL					
79	11	147	AARALAHGV					
79	11	147	AARALAHGVRV					
100	14	1264	AATLGFGA					
93	13	1264	AATLGFGAYM					
86	12	1187	AAVCTRGV					
79	11	1187	AAVCTRGVA					
79	11	1187	AAVCTRGVAKA					
93	13	1890	AILSPGAL					
86	12	1890	AILSPGALV	0.0014				
86	12	1890	AILSPGALVV	0.0035				
100	14	150	ALAHGVRV					
100	14	150	ALAHGVRVL	0.0037				
86	12	1737	ALGLLQTA					
86	12	669	ALSTGLIHL	0.0160	0.0008	0.2200	0.0002	0.0039
79	11	1696	ALVVGVVCA	0.0010				
79	11	1696	ALVVGVVCAA					
79	11	1696	ALVVGVVCAAI					
86	12	1602	AQAPPPSWDQM					
79	11	1251	AQGYKVLV					
79	11	1251	AQGYKVLVL					
86	12	77	AQPGYPWPL					
93	13	1265	ATLGFGAYM					
79	11	1354	ATPPGSVT					
79	11	1596	ATVCARAQA					
100	14	1419	AVAYYRGL					
100	14	1419	AVAYYRGLDV	0.0002				
79	11	1188	AVCTRGVA					
79	11	1188	AVCTRGVAKA					
79	11	1188	AVCTRGVAKAV					
100	14	1917	AVQWMNRL					
100	14	1917	AVQWMNRLI	0.0001				
100	14	1917	AVQWMNRLIA					
93	13	1903	CAAILRRHV					
79	11	1530	CAWYELTPA					
86	12	2941	CLRKLGVPPL	0.0002				
86	12	739	CLWMMLLI					
79	11	1653	CMSADLEV					

HCV A02 Super Motif with Binding Information

Conservancy	Freq.	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
79	11	1653	CMSADLEVV	0.0067				
79	11	1653	CMSADLEVVT					
79	11	1128	CTCGSSDL					
79	11	1128	CTCGSSDLYL					
79	11	1128	CTCGSSDLYLV					
79	11	1190	CTRGVAKA					
79	11	1190	CTRGVAKAV					
79	11	555	CTWMNSTGFT					
86	12	1462	CVTQTVDFSL	0.0006				
79	11	1527	DAGCAWYEL					
100	14	1574	DAHFLSQT					
86	12	1655	DILAGYGA					
79	11	1655	DILAGYGAGV	0.0002				
79	11	1655	DILAGYGAGVA					
86	12	279	DLCGSVFL					
79	11	279	DLCGSVFLV	0.0007				
86	12	1657	DLEVVTST					
86	12	1657	DLEVVTSTWV	0.0002				
86	12	1657	DLEVVTSTWVL					
~~93~~	~~13~~	~~2617~~	~~DLGVRVGEKM~~					
93	13	2617	DLGVRVCEKMA					
79	11	132	DLMGYIPL					
79	11	132	DLMGYIPLV	0.0630	0.0009	0.0490	0.0077	3.3000
79	11	132	DLMGYIPLVGA					
79	11	2412	DLSDGSWST					
79	11	2412	DLSDGSWSTV	0.0008				
79	11	1663	DLVNLLPA					
79	11	1663	DLVNLLPAI	0.0001				
79	11	1663	DLVNLLPAIL	0.0001				
79	11	2772	DLVVICESA					
86	12	1134	DLYLVTRHA	0.0001				
86	12	1134	DLYLVTRHADV					
86	12	321	DMMMNWSPT					
86	12	1339	DQAETAGA					
86	12	1339	DQAETAGARL					
86	12	1339	DQAETAGARLV					
86	12	994	DTAACGDI					
86	12	994	DTAACGDII					
86	12	124	DTLTCGFA					
86	12	124	DTLTCGFADL					
86	12	124	DTLTCGFADLM					
93	13	2673	DTRCFDST					

HCV A02 Super Motif with Binding Information

Conservancy	Freq.	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
93	13	2673	DTRCFDSTV					
93	13	2673	DTRCFDSTVT					
86	12	21	DVKFPGGGQI	0.0001				
86	12	21	DVKFPGGGQIV					
79	11	750	EAALENLV					
100	14	2794	EAMTRYSA					
86	12	2237	EANLLWRQEM					
93	13	1377	EIPFYGKA					
93	13	1377	EIPFYGKAI	0.0001				
100	14	2814	ELITSCSSNV	0.0002				
79	11	666	ELSPLLLST					
79	11	666	ELSPLLLSTT					
86	12	2245	EMGGNITRV	0.0003				
86	12	1731	EQFKQKAL					
86	12	1731	EQFKQKALGL					
86	12	1731	EQFKQKALGLL					
86	12	1342	ETAGARLV					
86	12	1342	ETAGARLVV					
86	12	1342	ETAGARLVVL					
86	12	1342	ETAGARLVVLA					
86	12	1207	ETTMRSPV					
86	12	1207	ETTMRSPVFT					
86	12	1659	EVVTSTWV					
86	12	1659	EVVTSTWVL	0.0001				
86	12	1659	EVVTSTWVLV	0.0004				
93	13	130	FADLMGYI					
79	11	130	FADLMGYIPL					
79	11	130	FADLMGYIPLV					
100	14	1927	FASRGNHV					
86	12	1927	FASRGNHVSPT					
100	14	1773	FISGIQYL					
100	14	1773	FISGIQYLA	0.1000				
100	14	1773	FISGIQYLAGL					
79	11	1304	FLADGGCSGGA					
86	12	177	FLLALLSCL	0.0046				
86	12	177	FLLALLSCLT					
93	13	728	FLLLADARV	0.2800	0.0480	0.0670	0.0150	0.3600
86	12	1228	FQVAHLHA					
86	12	1228	FQVAHLHAPT					
79	11	2646	FQYSPGQRV					
100	14	2792	FTEAMTRYSA					
93	13	1587	FTGLTHIDA					

HCV A02 Super Motif with Binding Information

Conservancy	Freq.	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
93	13	512	FTPSPVVV					
93	13	512	FTPSPVVVGT					
93	13	512	FTPSPVVVGTT					
79	11	684	FTTLPALST					
79	11	684	FTTLPALSTGL					
79	11	146	GAARALAHGV					
86	12	992	GADTAACGDI					
86	12	992	GADTAACGDII					
86	12	1861	GAGVAGAL					
86	12	1861	GAGVAGALV					
86	12	1861	GAGVAGALVA					
86	12	350	GAHWGVLA					
79	11	1895	GALVVGVV					
79	11	1895	GALVVGVVCA					
79	11	1895	GALVVGVVCAA					
86	12	1345	GARLVVLA					
79	11	1345	GARLVVLAT					
79	11	1345	GARLVVLATA					
79	11	1345	GARLVVLATAT					
100	14	1916	GAVQWMNRL	0.0001				
100	14	1916	GAVQWMNRLI					
100	14	1916	GAVQWMNRLIA					
100	14	1333	GIGTVLDQA					
100	14	1333	GIGTVLDQAET					
100	14	1776	GIQYLAGL					
100	14	1776	GIQYLAGLST					
100	14	1776	GIQYLAGLSTL					
79	11	1426	GLDVSVIPT					
93	13	1552	GLPVCQDHL	0.0001				
79	11	966	GLRDLAVA					
79	11	966	GLRDLAVAV	0.0034				
100	14	1782	GLSTLPGNPA					
79	11	1782	GLSTLPGNPAI					
93	13	1569	GLTHIDAHFL	0.0007				
93	13	28	GQVGGVYL					
93	13	28	GQVGGVYLL					
79	11	2063	GTFPINAYT					
79	11	2063	GTFPINAYTT					
100	14	1335	GTVLDQAET					
100	14	1335	GTVLDQAETA					
86	12	1863	GVAGALVA					
79	11	1081	GVCWTVYHGA					

HCV A02 Super Motif with Binding Information

Conservancy	Freq.	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
86	12	1670	GVLAALAA					
86	12	1670	GVLAALAAYCL					
79	11	161	GVNYATGNL	0.0001				
86	12	45	GVRATRKT					
100	14	2619	GVRVCEKM					
100	14	2619	GVRVCEKMA					
100	14	2619	GVRVCEKMAL	0.0002				
93	13	154	GVRVLEDGV	0.0001				
79	11	1900	GVVCAAIL					
100	14	1234	HAPTGSGKST					
100	14	1572	HIDAHFLSQT					
86	12	696	HLHQNIVDV	0.0100	0.0014	0.5400	0.0027	0.0037
79	11	1719	HLPYIEQGM					
93	13	1769	HMWNFISGI	0.3300	0.0004	0.1300	0.0280	0.0053
79	11	698	HQNIVDVQYL					
79	11	222	HTPGCVPCV					
86	12	2855	HTPVNSWL					
86	12	2855	HTPVNSWLGN					
79	11	1910	HVGPGEGA					
79	11	1910	HVGPGEGAV					
86	12	1933	HVSPTHYV					
100	14	1925	IAFASRGNHV					
79	11	1858	ILAGYGAGV	0.0430	0.0300	2.0000	0.0049	0.0450
79	11	1858	ILAGYGAGVA	0.0002				
86	12	1816	ILGGWVAA					
86	12	1816	ILGGWVAAQL	0.0430	0.0024	0.0190	0.0005	0.0039
86	12	1816	ILGGWVAAQLA					
86	12	1331	ILGIGTVL					
86	12	1331	ILGIGTVLDQA					
93	13	1891	ILSPGALV					
93	13	1891	ILSPGALVV	0.0210	0.0004	0.3700	0.0036	0.0130
93	13	1891	ILSPGALVVGV					
79	11	2591	IMAKNEVFCV	0.0088				
100	14	1777	IQYLAGLST					
100	14	1777	IQYLAGLSTL					
86	12	2250	ITRVESENKV					
86	12	2250	ITRVESENKVV					
100	14	2816	ITSCSSNV					
100	14	2816	ITSCSSNVSV					
100	14	2816	ITSCSSNVSVA					
86	12	909	ITWGADTA					
86	12	989	ITWGADTAA					

HCV A02 Super Motif with Binding Information

Conservancy	Freq.	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
79	11	1296	ITYSTYGKFL					
79	11	1296	ITYSTYGKFLA					
79	11	2613	IVFPDLGV					
79	11	2613	IVFPDLGVRV	0.0016				
93	13	30	IVGGVYLL					
86	12	1738	KALGLLQT					
86	12	1738	KALGLLQTA					
86	12	2625	KMALYDVV					
86	12	1734	KQKALGLL					
86	12	1734	KQKALGLLQT					
86	12	1734	KQKALGLLQTA					
86	12	121	KVIDTLTCGFA					
100	14	1255	KVLVLNPSV	0.0046				
100	14	1255	KVLVLNPSVA					
100	14	1255	KVLVLNPSVAA					
79	11	1244	KVPAAYAA					
86	12	1872	LAALAAYCL	0.0011				
79	11	1305	LADGGCSGGA					
86	12	1729	LAEQFKQKA					
86	12	1729	LAEQFKQKAL					
79	11	1657	LAGYGAGV					
79	11	1657	LAGYGAGVA					
79	11	1657	LAGYGAGVAGA					
100	14	151	LAHGVRVL					
86	12	179	LALLSCLT					
79	11	972	LAVAVEPV					
100	14	1924	LIAFASRGNHV					
100	14	2615	LITSCSSNV	0.0004				
100	14	2615	LITSCSSNVSV					
79	11	2612	LIVFPDLGV	0.0002				
79	11	2612	LIVFPDLGVRV					
86	12	178	LLALLSCL					
86	12	178	LLALLSCLT					
100	14	726	LLFLLLADA	0.0230	0.0150	0.0220	0.0011	0.0130
93	13	726	LLFLLLADARV					
86	12	1812	LLFNILGGWV	1.2000	0.0380	3.1000	0.1900	1.2000
86	12	1812	LLFNILGGWVA					
93	13	729	LLLADARV					
93	13	1887	LLPAILSPGA	0.0061				
93	13	1887	LLPAILSPGAL					
93	13	36	LLPRRGPRL	0.0025				
93	13	36	LLPRRGPRLGV					

HCV A02 Super Motif with Binding Information

Conservancy	Freq.	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
86	12	2240	LLWRQEMGGN					
93	13	1629	LLYRLGAV					
79	11	133	LMGYIPLV					
79	11	133	LMGYIPLVGA					
88	12	2761	LQDCTMLV					
88	12	126	LTCGFADL					
86	12	126	LTCGFADLM					
100	14	2180	LTDPSHIT					
100	14	2180	LTDPSHITA					
86	12	1052	LTGRDKNQV					
93	13	1570	LTHIDAHFL					
93	13	2176	LTSMLTDPSHI					
79	11	2738	LTTSCGNT					
79	11	2738	LTTSCGNTL					
79	11	2738	LTTSCGNTLT					
86	12	1591	LVAYQATV					
86	12	1591	LVAYQATVCA	0.0002				
79	11	1853	LVDILAGYGA	-0.0001				
86	12	1867	LVGGVLAA					
86	12	1867	LVGGVLAAL	0.0003				
86	12	1667	LVGGVLAALA					
86	12	1667	LVGGVLAALAA					
100	14	1257	LVLNPSVA					
100	14	1257	LVLNPSVAA					
100	14	1257	LVLNPSVAAT					
100	14	1257	LVLNPSVAATL					
79	11	1884	LVNLLPAI					
79	11	1884	LVNLLPAIL	0.0002				
86	12	1137	LVTRHADV					
79	11	1137	LVTRHADVI	0.0001				
79	11	1137	LVTRHADVIPV					
79	11	1897	LVVGVVCA					
79	11	1697	LVVGVVCAA					
79	11	1897	LVVGVVCAAI	0.0011				
79	11	1897	LVVGVVCAAIL					
79	11	2773	LVVICESA					
86	12	1348	LVVLATAT					
86	12	2582	MAKNEVFCV	0.0022				
100	14	2179	MLTDPSHI					
100	14	2179	MLTDPSHIT	0.0002				
100	14	2179	MLTDPSHITA					
93	13	322	MMMNWSPT					

HCV A02 Super Motif with Binding Information

Conservancy	Freq.	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
93	13	1418	NAVAYYRGL					
93	13	1418	NAVAYYRGLDV					
86	12	2066	NAYTTGPCT					
86	12	1815	NILGGWVA					
86	12	1815	NILGGWVAA					
86	12	1815	NILGGWVAAQL					
93	13	1282	NIRTGVRT					
79	11	1282	NIRTGVRTI	0.0001				
79	11	1282	NIRTGVRTIT					
79	11	1282	NIRTGVRTITT					
86	12	2249	NITRVESENKV					
86	12	700	NIVDVQYL					
86	12	116	NLGKVIDT					
86	12	116	NLGKVIDTL	0.0006				
86	12	116	NLGKVIDTLT					
93	13	1886	NLLPAILSPGA					
86	12	2239	NLLWRQEM					
93	13	166	NLPGCSFSI	0.0041				
93	13	166	NLPGCSFSIFL					
86	12	1460	NTCVTQTV					
93	13	416	NTNGSWHI					
86	12	14	NTNRRPQDV					
93	13	1889	PAILSPGA					
93	13	1889	PAILSPGAL					
86	12	1889	PAILSPGALV					
86	12	1889	PAILSPGALVV					
86	12	686	PALSTGLI					
86	12	686	PALSTGLIHL					
79	11	2609	PARLIVFPDL					
79	11	2066	PINAYTTGPCT					
79	11	1295	PITYSTYGKFL					
93	13	2403	PLEGEPGDPDL					
79	11	143	PLGGAARA					
79	11	143	PLGGAARAL	0.0001				
79	11	143	PLGGAARALA					
93	13	1628	PLLYRLGA					
93	13	1628	PLLYRLGAV	0.0001				
79	11	2667	PMGFSYDT					
79	11	2807	PQPEYDLEL					
79	11	2807	PQPEYDLELI					
79	11	2807	PQPEYDLELIT					
93	13	7	PQRKTKRNT					

1007 345 .072202

HCV A02 Super Motif with Binding Information

Conservancy	Freq.	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
86	12	109	PTDPRRRSRNL					
79	11	1473	PTFTIETT					
79	11	1473	PTFTIETTT					
100	14	1236	PTGSGKST					
93	13	1236	PTGSGKSTKV					
86	12	1936	PTHYVPESDA					
86	12	1936	PTHYVPESDAA					
79	11	1621	PTLHGPTPL					
79	11	1621	PTLHGPTPLL					
79	11	2070	PTLWARMI					
79	11	2870	PTLWARMIL					
79	11	2870	PTLWARMILM					
79	11	2870	PTLWARMILMT					
100	14	1626	PTPLLYRL					
93	13	1626	PTPLLYRLGA					
93	13	1626	PTPLLYRLGAV					
100	14	2857	PVNSWLGNI	0.0001				
100	14	2857	PVNSWLGNII	0.0001				
86	12	2857	PVNSWLGNIIM					
79	11	2318	PVVHGCPL					
93	13	508	PVYCFTPSPV	0.0004				
93	13	508	PVYCFTPSPVV					
86	12	1340	QAETAGARL					
86	12	1340	QAETAGARLV					
86	12	1340	QAETAGARLVV					
86	12	1603	QAPPPSWDQM					
93	13	1595	QATVCARA					
79	11	1595	QATVCARAQA					
93	13	29	QIVGGVYL					
93	13	29	QIVGGVYLL	0.0015				
86	12	336	QLLRIPQA					
86	12	2164	QLPCEPEPDV	0.0002				
79	11	2210	QLSAPSLKA					
79	11	2210	QLSAPSLKAT					
86	12	1465	QTVDFSLDPT					
86	12	1229	QVAHLHAPT					
86	12	1166	RAAVCTRGV					
79	11	1166	RAAVCTRGVA					
100	14	149	RALAHGVRV	0.0001				
100	14	149	RALAHGVRVL					
86	12	2733	RASGVLTT					
79	11	43	RLGVRATRKT					

HCV A02 Super Motif with Binding Information

Conservancy	Freq.	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
79	11	2918	RLHGLSAFSL	0.0280	0.0055	0.0180	0.0002	0.0032
79	11	2611	RLIVFPDL					
79	11	2611	RLIVFPDLGV	0.0890	0.0110	1.0000	0.0100	0.0050
79	11	1618	RLKPTLHGPT					
86	12	1029	RLLAPITA					
86	12	1347	RLVVLATA					
86	12	1347	RLVVLATAT					
100	14	619	RLWHYPCT					
86	12	317	RMAWDMMM					
93	13	635	RMYVGGVEHRL					
88	12	2243	RQEMGGNI					
88	12	2243	RQEMGGNIT					
86	12	2243	RQEMGGNITRV					
79	11	1284	RTGVRTIT					
79	11	1284	RTGVRTITT					
100	14	2621	RVCEKMAL					
86	12	2621	RVCEKMALYDV					
86	12	2252	RVESENKV					
86	12	2252	RVESENKVV	0.0001				
79	11	2100	RVGDFHYV					
86	12	156	RVLEDGVNYA					
88	12	156	RVLEDGVNYAT					
86	12	2833	RVYYLTRDPT					
79	11	1655	SADLEVVT					
79	11	1655	SADLEVVTST					
79	11	2212	SAPSLKAT					
79	11	2212	SAPSLKATCT					
93	13	2207	SASQLSAPSL					
100	14	175	SIFLLALL					
86	12	175	SIFLLALLSCL					
100	14	1470	SLDPTFTI					
86	12	1470	SLDPTFTIET					
79	11	1470	SLDPTFTIETT					
79	11	2926	SLHSYSPGEI	0.0008				
86	12	1051	SLTGRDKNQV	0.0002				
100	14	2178	SMLTDPSHI	0.0053				
100	14	2178	SMLTDPSHIT					
100	14	2178	SMLTDPSHITA					
86	12	2163	SQLPCEPEPDV					
93	13	2209	SQLSAPSL					
79	11	2209	SQLSAPSLKA					
79	11	2209	SQLSAPSLKAT					

HCV A02 Super Motif with Binding Information

Conservancy	Freq.	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
93	13	56	SQPRGRRQPI					
86	12	1242	STKVPAAYA					
79	11	1242	STKVPAAYAA					
100	14	1784	STLPGNPA					
79	11	1784	STLPGNPAI	0.0007				
79	11	2	STNPKPQRKT					
86	12	1663	STWVLVGGV					
86	12	1663	STWVLVGGVL					
86	12	1663	STWVLVGGVLA					
86	12	1299	STYGKFLA					
100	14	1262	SVAATLGFGA					
86	12	1455	SVIDCNTCV	0.0088				
86	12	1455	SVIDCNTCVT					
86	12	995	TAACGDII					
86	12	1343	TAGARLVV					
86	12	1343	TAGARLVVL					
86	12	1343	TAGARLVVLA					
79	11	1343	TAGARLVVLAT					
79	11	2852	TARHTPVNSWL					
79	11	2590	TIMAKNEV					
93	13	1266	TLGFGAYM					
86	12	1266	TLGFGAYMSKA					
79	11	1622	TLHGPTPL					
79	11	1622	TLHGPTPLL	0.0070				
86	12	1811	TLLFNILGGWV					
79	11	686	TLPALSTGL	0.0003				
79	11	686	TLPALSTGLI	0.0004				
79	11	1785	TLPGNPAI					
86	12	125	TLTCGFADL	0.0003				
86	12	125	TLTCGFADLM					
79	11	2871	TLWARMIL					
79	11	2871	TLWARMILM					
79	11	2871	TLWARMILMT					
86	12	1209	TMRSPVFT					
86	12	1464	TQTVDFSL					
86	12	1464	TQTVDFSLDPT					
79	11	2589	TTIMAKNEV					
79	11	685	TTLPALST					
79	11	685	TTLPALSTGL					
79	11	685	TTLPALSTGLI					
86	12	1208	TTMRSPVFT					
79	11	2739	TTSCGNTL					

HCV A02 Super Motif with Binding Information

Conservancy	Freq.	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
79	11	2739	TTSCGNTLT					
79	11	1597	TVCARAQA					
86	12	1466	TVDFSLDPT					
86	12	1466	TVDFSLDPTFT					
100	14	1336	TVLDQAET					
100	14	1336	TVLDQAETA					
86	12	1336	TVLDQAETAGA					
100	14	1263	VAATLGFGA					
93	13	1263	VAATLGFGAYM					
86	12	1230	VAHLHAPT					
86	12	1440	VATDALMT					
86	12	1592	VAYQATVCA	0.0005				
79	11	1592	VAYQATVCARA					
100	14	1420	VAYYRGLDV	0.0001				
100	14	1420	VAYYRGLDVSV					
86	12	1456	VIDCNTCV					
86	12	1456	VIDCNTCVT					
86	12	1456	VIDCNTCVTQT					
86	12	122	VIDTLTCGFA					
86	12	1671	VLAALAAYCL	0.0500	0.0087	0.0047	0.0002	0.0550
93	13	1521	VLCECYDA					
79	11	1521	VLCECYDAGCA					
100	14	1337	VLDQAETA					
86	12	1337	VLDQAETAGA					
86	12	157	VLEDGVNYA					
86	12	157	VLEDGVNYAT					
100	14	1258	VLNPSVAA					
100	14	1258	VLNPSVAAT					
100	14	1258	VLNPSVAATL	0.0015				
79	11	2737	VLTTSCGNT					
79	11	2737	VLTTSCGNTL	0.0002				
79	11	2737	VLTTSCGNTLT					
79	11	1852	VLVDILAGYGA					
86	12	1666	VLVGGVLA					
86	12	1666	VLVGGVLAA	0.0270	0.0130	0.3100	0.0120	0.0130
86	12	1666	VLVGGVLAAL	0.0084				
86	12	1666	VLVGGVLAALA					
100	14	1256	VLVLNPSV					
100	14	1256	VLVLNPSVA	0.0009				
100	14	1256	VLVLNPSVAA					
100	14	1256	VLVLNPSVAAT					
79	11	2800	VQPEKGGRKPA					

<u>HCV A02 Super Motif with Binding Information</u>

Conservancy	Freq.	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
100	14	1918	VQWMNRLI					
100	14	1918	VQWMNRLIA					
100	14	1918	VQWMNRLIAFA					
86	12	1463	VTQTVDFSL					
79	11	1138	VTRHADVI					
79	11	1138	VTRHADVIPV					
86	12	1661	VTSTWVLV					
86	12	1661	VTSTWVLVGGV					
79	11	1439	VVATDALM					
79	11	1439	VVATDALMT					
79	11	1901	VVCAAILRRHV					
79	11	1898	VVGVVCAA					
79	11	1898	VVGVVCAAI					
79	11	1898	VVGVVCAAIL					
86	12	1660	VVTSTWVL					
86	12	1660	VVTSTWVLV	0.0003				
86	12	1766	WAKHMWNFI	0.0001				
86	12	76	WAQPGYPWPL					
86	12	2873	WARMILMT					
79	11	2297	WARPDYNPPL					
100	14	1920	WMNRLIAFA	0.0410	0.0330	3.0000	0.0023	0.1000
79	11	557	WMNSTGFT					
86	12	1665	WVLVGGVL					
86	12	1665	WVLVGGVLA	0.0005				
86	12	1665	WVLVGGVLAA	0.0015				
86	12	1665	WVLVGGVLAAL					
79	11	1249	YAAQGYKV					
79	11	1249	YAAQGYKVL					
79	11	1249	YAAQGYKVLV					
79	11	1249	YAAQGYKVLVL					
79	11	136	YIPLVGAPL	0.0050				
100	14	1779	YLAGLSTL					
86	12	1165	YLKGSSGGPL	0.0002				
86	12	1165	YLKGSSGGPLL					
93	13	35	YLLPRRGPRL	0.0400	0.0007	0.0220	0.0089	0.0039
79	11	2836	YLTRDPTT					
86	12	1590	YLVAYQAT					
86	12	1590	YLVAYQATV	0.2500	0.1100	0.6300	0.0450	1.2000
86	12	1590	YLVAYQATVCA					
86	12	1136	YLVTRHADV	0.0110	0.0021	2.8000	0.0520	0.0130
79	11	1136	YLVTRHADVI					
93	13	1594	YQATVCARA					

HCV A02 Super Motif with Binding Information

Conservancy	Freq.	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
79	11	1594	YQATVCARAQA					
79	11	1106	YTNVDQDL					
79	11	1106	YTNVDQDLV					
86	12	276	YVGDLCGSV	0.0018				
86	12	276	YVGDLCGSVFL					
93	13	637	YVGGVEHRL	0.0008				
86	12	1939	YVPESDAA					
86	12	1939	YVPESDAAA					
86	12	1939	YVPESDAAARV					
			555					

Table IX

HCV A03 Super Motif (With Binding Information)

Conservancy	Freq.	Position	Sequence	A*0301	A*1101	A*3101	A*3301	A*6801
86	12	647	AACNWTRGER	0.0003	0.0140	0.0450	0.0055	0.0018
79	11	147	AARALAHGVR					
79	11	1167	AAVCTRGVAK					
79	11	2208	ASQLSAPSLK					
86	12	1265	ATLGFGAYMSK					
79	11	48	ATRKTSER					
79	11	1168	AVCTRGVAK	0.0260	0.0250	0.0011	0.0004	0.0001
86	12	2941	CLRKLGVPPLR					
79	11	565	CTWMNSTGFTK	0.7600	0.7500			
79	11	2599	CVQPEKGGR	0.0008	0.0005			
79	11	2599	CVQPEKGGRK	0.0011	0.0008			
100	14	1574	DAHFLSQTK	0.0003	0.0005			
93	13	2617	DLGVRVCEK	0.0003	0.0002	0.0008	0.0440	0.0002
79	11	1143	DVIPVRRR					
86	12	2245	EMGGNITR					
86	12	2598	EVRCVQPEK	0.0008	0.0270	0.0003	0.0006	0.4500
100	14	728	FLLLADAR					
79	11	146	GAARALAHGVR					
100	14	1918	GAVQWMNR					
79	11	3037	GIYLLPNR					
79	11	1004	GLPVSARR					
86	12	1131	GSSDLYLVTR					
86	12	1883	GVAGALVAFK	0.3900	1.4000	0.0055	0.0011	0.0680
79	11	3036	GVQIYLLPNR	0.0014	0.0140	0.1500	0.0130	0.0007
79	11	45	GVRATRKTSER					
79	11	1900	GVVCAAILR					
79	11	1900	GVVCAAILRR					
93	13	33	GVYLLPRR					
93	13	33	GVYLLPRRGPR					
79	11	1141	HADVIPVR					
79	11	1141	HADVIPVRR					
79	11	1141	HADVIPVRRR					
100	14	1234	HAPTGSGK					
93	13	1234	HAPTGSGKSTK					
100	14	1572	HIDAHFLSQTK					
86	12	1232	HLHAPTGSGK	0.5900	0.0024	0.0005	0.0006	0.0028
100	14	1395	HLIFCHSK					
100	14	1395	HLIFCHSKK	0.0250	0.0006	0.0003	0.0004	0.0010
100	14	1395	HLIFCHSKKK	0.0260	0.0002	0.0009	0.0006	0.0001
79	11	2920	HSYSPGEINR					
79	11	222	HTPGCVPCVR	0.0004	0.0012			
86	12	2250	ITRVESENK	0.0150	0.0079	0.0007	0.0006	0.0092
86	12	1298	ITYSTYGK					
79	11	2813	IVFPDLGVR	0.0036	0.0044			
93	13	30	IVGGVYLLPR	0.0008	0.0056			
93	13	30	IVGGVYLLPRR					
86	12	2944	KLGVPPLR					
86	12	10	KTKRNTNR					
86	12	10	KTKRNTNRR	0.0110	0.0100			
93	13	51	KTSERSQPR	0.1600	0.0640	0.2700	0.0160	0.0550
86	12	51	KTSERSQPRGR					
86	12	1729	LAEQFKQK					

HCV A03 Super Motif (With Binding Information)

Conservancy	Freq.	Position	Sequence	A*0301	A*1101	A*3101	A*3301	A*6801
86	12	2235	LIEANLLWR	0.0008	0.0005	0.0018	0.0069	0.0008
100	14	1396	LIFCHSKK					
100	14	1398	LIFCHSKKK	0.5400	0.1900	0.0071	0.0012	0.0240
79	11	2812	LVFPDLGVR	0.0003	0.0001			
100	14	726	LLFLLLADAR					
93	13	36	LLPRRGPR					
86	12	97	LLSPRGSR					
79	11	1591	LVAYQATVCAR					
79	11	1	MSTNPKPQR					
79	11	1	MSTNPKPQRK					
86	12	2249	NITRVESENK	0.0010	0.0062			
79	11	14	NTNRRPQDVK	0.0010	0.0007			
79	11	1295	PITYSTYGK					
79	11	2667	PMGFSYDTR					
93	13	514	PSPVVVGTTDR					
79	11	1607	PSWDQMWK					
86	12	109	PTQPRRRSR	0.0008	0.0005			
93	13	1236	PTQSGKSTK	0.0002	0.0001	0.0008	0.0006	0.0002
93	13	616	PVVVGTTDR	0.0008	0.0006			
86	12	1340	QAETAGAR					
93	13	29	QIVGGVYLLPR					
86	12	289	QLFTFSPR					
79	11	289	QLFTFSPRR	0.7500	0.0330	0.0290	0.0077	3.1000
79	11	2210	QLSAPSLK					
79	11	1186	RAAVCTRGVAK					
100	14	149	RALAHGVR					
79	11	47	RATRKTSER					
79	11	43	RLGVRATR					
79	11	43	RLGVRATRK	0.9400	0.0280	0.0420	0.0004	0.0001
100	14	1923	RLIAFASR					
79	11	2811	RLVFPDLGVR					
100	14	635	RMYVGGVEHR	0.7200	0.0200	0.1900	0.0030	0.0045
93	13	55	RSQPRGRR					
79	11	2207	SASQLSAPSLK					
86	12	1132	SSDLYLVTR	0.0003	0.0044			
79	11	2	STNPKPQR					
79	11	2	STNPKPQRK					
79	11	2	STNPKPQRKTK					
86	12	1266	TLGFGAYMSK	0.0810	0.0610	0.0005	0.0013	0.0009
79	11	1622	TLHGPTPLLYR					
93	13	52	TSERSQPR					
86	12	52	TSERSQPRGR	0.0003	0.0001			
86	12	52	TSERSQPRGRR					
86	12	1050	TSLTGRDK					
86	12	1664	VAGALVAFK	0.2400	0.8900	0.0048	0.0025	0.0310
79	11	1592	VAYQATVCAR	0.0005	0.0038	0.0680	0.0720	0.0280
86	12	1337	VLDQAETAGAR					
79	11	1136	VTRHADVIPVR					
79	11	1901	VVCAAILR					
79	11	1901	VVCAAILRR					
79	11	1898	VVGVVCAAILR					
93	13	517	VVVGTTDR					

HCV A03 Super Motif (With Binding Information)

Conservancy	Freq.	Position	Sequence	A*0301	A*1101	A*3101	A*3301	A*6801
88	12	93	WAGWLLSPR					
88	12	96	WLLSPRGSR	0.0008	0.0005			
100	14	1920	WMNRLIAFASR					
79	11	557	WMNSTGFTK	0.0530	0.0810	0.0014	0.0420	0.0058
93	13	35	YLLPRRGPR	0.0054	0.0005			
79	11	2930	YSPGEINR					
100	14	637	YVGGVEHR					
86	12	1939	YVPESDAAAI	0.0003	0.0001			
			112					

Table X **HCV A24 Super Motif With Binding Information**

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*2401
AILSPGAL	1890	8	13	93	
ALAHGVRVL	150	9	14	100	
ALSTGLIHL	689	9	12	86	
ALVVGVVCAAI	1896	11	11	79	
ATGNLPGCSF	165	10	13	93	
ATLGFGAY	1265	8	14	100	
ATLGFGAYM	1265	9	13	93	
AVAYYRGL	1419	8	14	100	
AVQWMNRL	1917	8	14	100	
AVQWMNRLI	1917	9	14	100	
AVQWMNRLIAF	1917	11	14	100	
AWDMMMNW	319	8	12	86	
AYAAQGYKVL	1248	10	11	79	0.0009
AYYRGLDVSVI	1421	11	14	100	
CLRKLGVPPL	2941	10	12	86	
CLWMMLLI	739	8	12	86	
CTCGSSDL	1128	8	11	79	
CTCGSSDLY	1128	9	11	79	0.0001
CTCGSSDLYL	1128	10	11	79	
CTRGVAKAVDF	1190	11	11	79	
CTWMNSTGF	555	9	11	79	
CVTQTVDF	1462	8	12	86	
CVTQTVDFSL	1462	10	12	86	
CYDAGCAW	1525	8	11	79	
CYDAGCAWY	1525	9	11	79	
CYDAGCAWYEL	1525	11	11	79	
DFSLDPTF	1468	8	14	100	
DFSLDPTFTI	1468	10	14	100	
DLCGSVFL	279	8	12	86	
DLEVVTSTW	1657	9	12	86	
DLEVVTSTWVL	1657	11	12	86	
DLGVRVCEKM	2617	10	13	93	
DLMGYIPL	132	8	11	79	
DLVNLLPAI	1883	9	11	79	
DLVNLLPAIL	1883	10	11	79	
DTAACGDI	994	8	12	86	
DTAACGDII	994	9	12	86	
DTLTCGFADL	124	10	12	86	
DTLTCGFADLM	124	11	12	86	
DVKFPGGGQI	21	10	12	86	
DYPYRLWHY	615	9	14	100	
EIPFYGKAI	1377	9	13	93	
ETAGARLVVL	1342	10	12	86	
ETTMRSPVF	1207	9	12	86	
EVVTSTWVL	1659	9	12	86	

HCV A24 Super Motif With Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*2401
FISGIQYL	1773	8	14	100	
FISGIQYLAGL	1773	11	14	100	
FLLALLSCL	177	9	12	86	
FTEAMTRY	2792	8	14	100	
FTGLTHIDAHF	1567	11	13	93	
FTTLPALSTGL	684	11	11	79	
FWAKHMWNF	1765	9	12	86	6.9000
FWAKHMWNFI	1765	10	12	86	
GFADLMGY	129	8	13	93	
GFADLMGYI	129	9	13	93	
GFADLMGYIPL	129	11	11	79	
GFSYDTRCF	2669	9	11	79	
GIQYLAGL	1776	8	14	100	
GIQYLAGLSTL	1776	11	14	100	
GLPVCQDHL	1652	0	13	93	
GLPVCQDHLEF	1552	11	12	06	
GLSAFSLHSY	2921	10	11	79	0.0001
GLSTLPGNPAI	1782	11	11	70	
GLTHIDAHF	1569	9	13	93	
GLTHIDAHFL	1569	10	13	93	
GTFPINAY	2063	8	11	79	
GVAGALVAF	1863	9	12	86	
GVAKAVDF	1193	8	11	79	
GVLAALAAY	1670	9	12	86	
GVLAALAAYCL	1670	11	12	86	
GVNYATGNL	161	9	11	79	
GVRVCEKM	2619	8	14	100	
GVRVCEKMAL	2619	10	14	100	
GVRVCEKMALY	2619	11	14	100	
GVRVLEDGVNY	154	11	12	86	
GVVCAAIL	1900	8	11	79	
GWRLLAPI	1027	8	11	79	
GWRLLAPITAY	1027	11	11	79	
GYGAGVAGAL	1859	10	12	86	0.0003
GYIPLVGAPL	135	10	11	79	0.0057
GYRRCRASGVL	2728	11	12	86	
HLHQNIVDVQY	698	11	11	79	
HLPYIEQGM	1719	9	11	79	
HMWNFISGI	1769	9	13	93	
HMWNFISGIQY	1769	11	13	93	
HTPVNSWL	2855	8	12	86	
HTPVNSWLGNI	2855	11	12	86	
HVGPGEGAVQW	1910	11	11	79	
IFLLALLSCL	176	10	12	86	
ILGGWVAAQL	1816	10	12	86	0.0026

HCV A24 Super Motif With Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*2401
ILGIGTVL	1331	8	12	86	
IMAKNEVF	2591	8	12	86	
ITYSTYGKF	1296	9	12	86	
ITYSTYGKFL	1296	10	11	79	
IVDVQYLY	701	8	12	86	
IVGGVYLL	30	8	13	93	
KFPGGGQI	23	8	13	93	
KVIDTLTCGF	121	10	12	86	
LFNILGGW	1813	8	12	86	
LIEANLLW	2235	8	12	86	
LINTNGSW	414	8	11	79	
LLALLSCL	170	0	12	86	
LLAPITAY	1030	8	14	00	
LLFNILGGW	1812	9	12	86	
LLPAILSPGAL	1887	11	13	93	
LLPRRGPRL	36	9	13	93	
LLSPRGSRPSW	97	11	11	79	
LLWRQEMGGNI	2240	11	12	86	
LTCGFADL	126	8	12	86	
LTCGFADLM	126	9	12	86	
LTCGFADLMGY	126	11	12	86	
LTHIDAHF	1570	8	13	93	
LTHIDAHFL	1570	9	13	93	
LTSMLTDPSHI	2178	11	13	93	
LTTSCGNTL	2738	9	11	79	
LVDILAGY	1853	8	11	79	
LVGGVLAAL	1667	9	12	86	
LVLNPSVAATL	1257	11	14	100	
LVNLLPAI	1804	8	11	79	
LVNLLPAIL	1884	9	11	79	
LVTRHADVI	1137	9	11	79	
LVVGVVCAAI	1897	10	11	79	
LVVGVVCAAIL	1897	11	11	79	
LWARMILM	2872	8	12	86	
LWARMILMTHF	2872	11	12	86	
LWRQEMGGNI	2241	10	12	86	
LYLVTRHADVI	1135	11	11	79	
MILMTHFF	2876	8	12	86	
MLTDPSHI	2179	8	14	100	
MWNFISGI	1770	8	14	100	
MWNFISGIQY	1770	10	14	100	
MWNFISGIQYL	1770	11	14	100	
MYVGGVEHRL	636	10	13	93	0.0270
NFISGIQY	1772	8	14	100	
NFISGIQYL	1772	9	14	100	0.0170

HCV A24 Super Motif With Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*2401
NILGGWVAAQL	1815	11	12	86	
NIRTGVRTI	1282	9	11	79	
NIVDVQYL	700	8	12	86	
NIVDVQYLY	700	9	12	86	0.0001
NLGKVIDTL	118	9	12	86	
NLLWRQEM	2239	8	12	86	
NLPGCSFSI	168	9	13	93	
NLPGCSFSIF	168	10	13	93	
NLPGCSFSIFL	168	11	13	93	
NTCVTQTVDF	1460	10	12	86	
NTNGSWHI	416	8	13	93	
NTNRRPQDVKF	14	11	11	79	
NVDQDLVGW	1108	9	11	79	
NWFGCTWM	551	8	12	86	
PITYSTYGKF	1295	10	11	79	
PITYSTYGKFL	1295	11	11	79	
PLEGEPGDPDL	2403	11	13	93	
PLGGAARAL	143	9	11	79	
PMGFSYDTRCF	2667	11	11	79	
PTDPRRRSRNL	109	11	12	86	
PTLHGPTPL	1621	9	11	79	
PTLHGPTPLL	1621	10	11	79	
PTLHGPTPLLY	1621	11	11	79	
PTLWARMI	2870	8	11	79	
PTLWARMIL	2870	9	11	79	
PTLWARMILM	2870	10	11	79	
PTPLLYRL	1626	8	14	100	
PVQDHLEF	1554	9	12	86	
PVQDHLEFW	1554	10	12	86	
PVNSWLGNI	2857	9	14	100	
PVNSWLGNII	2857	10	14	100	
PVNSWLGNIIM	2857	11	12	86	
PVVHGCPL	2318	8	11	79	
QFKQKALGL	1732	9	12	86	
QFKQKALGLL	1732	10	12	86	
QIVGGVYL	29	8	13	93	
QIVGGVYLL	29	9	13	93	
QTVDFSLDPTF	1465	11	12	86	
QWMNRLIAF	1919	9	14	100	
QYLAGLSTL	1778	9	14	100	0.0480
QYSPGQRVEF	2647	10	11	79	0.0180
QYSPGQRVEFL	2647	11	11	79	
RLHGLSAF	2918	8	12	86	
RLHGLSAFSL	2918	10	11	79	0.0001
RLIVFPDL	2611	8	11	79	

HCV A24 Super Motif With Binding Information

Sequence	Position	Peptide No.	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*2401
RLLAPITAY	1029		9	12	86	
RMAWDMMM	317		8	12	86	
RMAWDMMMNW	317		10	12	86	
RMILMTHF	2875		8	12	86	
RMILMTHFF	2875		9	12	86	
RMYVGGVEHRL	635		11	13	93	
RVCEKMAL	2621		8	14	100	
RVCEKMALY	2621		9	14	100	
RVLEDGVNY	156		9	12	86	
SFSIFLLAL	173		9	14	100	
SFSIFLLALL	173		10	14	100	0.0041
SIFLLALL	175		8	14	100	
SIFLLALLSCL	175		11	12	86	
SLDPTFTI	1470		8	14	100	
SLHSYSPGEI	2928		10	11	79	
SMLTDPSHI	2178		9	14	100	
STKVPAAY	1242		8	12	86	
STLPGNPAI	1784		9	11	79	
STWVLVGGVL	1663		10	12	86	
SVAATLGF	1262		8	14	100	
SVAATLGFGAY	1262		11	14	100	
SWDQMWKCL	1608		9	11	79	
SWLGNIIM	2860		8	12	86	
SYLKGSSGGPL	1164		11	12	86	
TIMAKNEVF	2590		9	11	79	
TLGFGAYM	1266		8	13	93	
TLHGPTPL	1622		8	11	79	
TLHGPTPLL	1622		9	11	79	
TLHGPTPLLY	1622		10	11	79	0.0001
TLLFNILGGW	1811		10	12	86	
TLPALSTGL	686		9	11	79	
TLPALSTGLI	686		10	11	79	
TLPGNPAI	1785		8	11	79	
TLTCGFADL	125		9	12	86	
TLTCGFADLM	125		10	12	86	
TLWARMIL	2871		8	11	79	
TLWARMILM	2871		9	11	79	
TTIMAKNEVF	2589		10	11	79	
TTLPALSTGL	685		10	11	79	
TTLPALSTGLI	685		11	11	79	
TTMRSPVF	1208		8	12	86	
TTSCGNTL	2739		8	11	79	
TVDFSLDPTF	1466		10	12	86	
TWMNSTGF	556		8	11	79	
TWVLVGGVL	1664		9	12	86	

HCV A24 Super Motif With Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*2401
TYSTYGKF	1297	8	13	93	
TYSTYGKFL	1297	9	12	86	0.0230
VFTGLTHI	1566	8	13	93	
VIDTLTCGF	122	9	12	86	
VLAALAAY	1671	8	12	86	
VLAALAAYCL	1671	10	12	86	0.0070
VLEDGVNY	157	8	12	86	
VLNPSVAATL	1258	10	14	100	
VLTTSCGNTL	2737	10	11	79	
VLVDILAGY	1852	9	11	79	
VLVGGVLAAL	1666	10	12	86	
VMGSSYGF	2639	8	11	79	
VMGSSYGFQY	2639	10	11	79	
VTQTVDFSL	1463	9	12	86	
VTRHADVI	1138	8	11	79	
VVATDALM	1439	8	11	79	
VVGVVCAAI	1898	9	11	79	
VVGVVCAAIL	1898	10	11	79	
VVTSTWVL	1660	8	12	86	
VYLLPRRGPRL	34	11	13	93	0.0016
WMNRLIAF	1920	8	14	100	
WVLVGGVL	1665	8	12	86	
WVLVGGVLAAL	1665	11	12	86	
YIPLVGAPL	136	9	11	79	
YLAGLSTL	1779	8	14	100	
YLKGSSGGPL	1165	10	12	86	
YLKGSSGGPLL	1165	11	12	86	
YILPRRGPRL	35	10	13	93	0.0001
YLVTRHADVI	1136	10	11	79	
YTNVDQDL	1108	8	11	79	
YTNVDQDLVGW	1108	11	11	79	
YVGDLCGSVF	276	10	12	86	
YVGDLCGSVFL	276	11	12	86	
YVGGVEHRL	637	9	13	93	
YYRGLDVSVI	1422	10	14	100	
260		3			

Table XI

HCV B07 Super Motif (with Binding Information)

Conservancy	Freq.	Position	Sequence	B*0702	B*3501	B*5101	B*5301	B*5401
86	12	1604	APPPSWDQM	0.0028	0.0002	0.0002	0.0001	0.0002
79	11	1604	APPPSWDQMW	0.0001	0.0001	0.0002	0.0006	0.0003
93	13	1235	APTGSGKSTKV	0.0001				
79	11	2869	APTLWARM	0.4300	0.0001	0.0012	-0.0002	0.0023
79	11	2869	APTLWARMI	0.0160	0.0002	0.0012	0.0001	0.0002
79	11	2869	APTLWARMIL	0.8000	0.0001	0.0010	0.0001	0.0003
79	11	2869	APTLWARMILM	0.0130	0.0001	-0.0003	-0.0002	0.0033
79	11	2410	DPDLSDGSW	0.0001	0.0002	0.0002	0.0005	0.0002
86	12	111	DPRRRSRNL	0.0170	0.0002	0.0001	0.0001	0.0002
79	11	2615	FPDLGVRV	0.0001				
100	14	24	FPGGGQIV	0.0001				
100	14	24	FPGGGQIVGGV	0.0001				
86	12	1912	GPGEGAVQW	0.0001	0.0002	0.0002	0.0001	0.0002
86	12	1912	GPGEGAVQWM	0.0001	0.0001	0.0002	0.0001	0.0003
93	13	41	GPRLGVRA	0.0001				
100	14	1625	GPTPLLYRL	0.0024	0.0002	0.0002	0.0001	0.0002
93	13	1625	GPTPLLYRLGA	0.0005				
93	13	507	GPVYCFTPSPV	0.0001				
93	13	1378	IPFYGKAI	0.0120	0.0001	0.1200	-0.0002	0.2000
79	11	137	IPLVGAPL	0.4400	0.0032	0.0700	0.0003	0.0035
86	12	2606	KPARLIVF	0.0150	0.0002	0.0017	-0.0002	0.0008
79	11	2606	KPARLIVFPDL	0.0003				
79	11	1820	KPTLHGPTPL	1.4150	0.0001	0.0002	0.0001	0.0003
79	11	1620	KPTLHGPTPLL	0.0021				
93	13	1888	LPAILSPGA	0.0001	0.0001	0.0001	0.0002	0.9400
93	13	1888	LPAILSPGAL	0.0053	0.0001	0.0036	0.0001	0.2100
86	12	1888	LPAILSPGALV	0.0003				
100	14	887	LPALSTGL	0.0020				
86	12	887	LPALSTGLI	0.0350	0.0002	2.0000	0.0082	0.0005
86	12	887	LPALSTGLIHL	0.0011				
86	12	2165	LPCEPEPDV	0.0001	0.0002	0.0001	0.0001	0.0002
93	13	169	LPGCSFSI	0.0110	0.0360	0.0059	0.0150	0.0018
93	13	169	LPGCSFSIF	0.1950	0.0790	0.0550	0.0013	0.0015
93	13	169	LPGCSFSIFL	0.0022	0.0009	0.0100	0.0140	0.0012
93	13	169	LPGCSFSIFLL	0.0007				
93	13	37	LPRRGPRL	6.5000	0.0001	0.0180	-0.0002	0.0020
93	13	37	LPRRGPRLGV	0.1900	0.0001	0.0009	0.0001	0.0025
93	13	1553	LPVCQDHL	0.0005				
86	12	1553	LPVCQDHLEF	0.0001	0.0048	0.0002	0.0110	0.0003
86	12	1553	LPVCQDHLEFW	0.0001				
86	12	1720	LPYIEQGM	0.0130	0.0001	0.0040	-0.0002	0.0013
100	14	1260	NPSVAATL	0.0011				
100	14	1260	NPSVAATLGF	0.0001	0.0001	0.0002	0.0001	0.0003
86	12	1605	PPPSWDQM	0.0003				
79	11	1605	PPPSWDQMW	0.0001	0.0002	0.0001	0.0001	0.0002
79	11	1606	PPSWDQMW	0.0002				
79	11	1606	PPSWDQMWKC	0.0001				
79	11	2317	PPVHGCPL	0.0140	0.0001	0.0001	0.0001	-0.0002
79	11	2601	QPEKGGRKPA	0.0011	0.0001	0.0001	0.0002	0.0190
79	11	2808	QPEYDLEL	0.0002				
79	11	2808	QPEYDLELI	0.0001	0.0002	0.0002	0.0001	0.0002
86	12	78	QPGYPWPL	0.0006				

HCV B07 Super Motif (with Binding Information)

Conservancy	Freq.	Position	Sequence	B*0702	B*3501	B*5101	B*5301	B*5401
86	12	78	OPGYPWPLY	0.0001	0.0011	0.0002	0.0001	0.0002
93	13	57	OPRGRROPI	0.2300	0.0002	0.0001	0.0001	0.0002
79	11	2299	RPDYNPPL	0.0050				
93	13	1893	SPGALVVGV	0.0001	0.0002	0.0002	0.1200	0.0002
79	11	1893	SPGALVVGVV	0.0130	0.0001	0.0018	0.0001	0.0003
79	11	2931	SPGEINRV	0.0007				
79	11	2931	SPGEINRVA	0.0003	0.0001	0.0001	0.0002	0.0037
79	11	2649	SPGQRVEF	0.0027				
79	11	2649	SPGQRVEFL	0.1200	0.0002	0.0002	0.0001	0.0002
79	11	99	SPRGSRPSW	0.3800	0.0002	0.0005	0.0001	0.0002
86	12	1935	SPTHYVPESDA	0.0001				
86	12	1975	TPCSGSWL	0.0028				
79	11	1126	TPCTCGSSDL	0.0005	0.0001	0.0002	0.0001	0.0003
79	11	1126	TPCTCGSSDLY	0.0001				
86	12	223	TPGCVPCV	0.0001				
93	13	1550	TPGLPVCQDHL	0.0001				
93	13	1627	TPLLYRLGA	0.0083	0.0001	0.0001	0.0002	0.2300
93	13	1627	TPLLYRLGAV	0.0120	0.0001	0.0008	0.0001	0.0110
86	12	2858	TPVNSWLGNI	0.0001	0.0001	0.0053	0.0008	0.0003
86	12	2858	TPVNSWLGNII	0.0001				
86	12	1940	VPESDAAA	0.0022				
86	12	1940	VPESDAAARV	0.0001	0.0001	0.0010	0.0001	0.0003
86	12	799	WPLLLLLL	0.0021				
100	14	616	YPYRLWHY	0.0001				

76

Table XII **HCV B27 Super Motif**

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
AKHMWNFI	1767	8	12	86
AKNEVFCV	2593	8	12	86
ARALAHGV	148	8	14	100
DRSELSPL	663	8	11	79
EKGGRKPA	2603	8	11	79
EKMALYDV	2624	8	12	86
FKQKALGL	1733	8	12	86
GHRMAWDM	315	8	13	93
GKSTKVPA	1240	8	12	86
GRKPARLI	2606	8	11	79
HRMAWDMM	316	8	13	93
IKGGRHLI	1390	8	11	79
IRTGVRTI	1283	8	11	79
KKCDELAA	1403	8	14	100
KKKCDELA	1402	8	14	100
LHGPTPLL	1623	8	11	79
LHQNIVDV	697	8	12	86
LRDLAVAV	969	8	11	79
NHVSPTHY	1932	8	12	86
PRGRRQPI	58	8	13	93
PRGSRPSW	100	8	11	79
PRRRSRNL	112	8	12	86
RHADVIPV	1140	8	11	79
RHTPVNSW	2854	8	12	86
RKLGVPPL	2943	8	12	86
RKPARLIV	2607	8	11	79
RRCRASGV	2730	8	13	93
RRGPRLGV	39	8	13	93
RRPQDVKF	17	8	12	86
SKKKCDEL	1401	8	14	100
SRNLGKVI	118	8	12	86
THIDAHFL	1571	8	13	93
TKLKLTPI	2985	8	12	86
TKVPAAYA	1243	8	12	86
TRCFDSTV	2674	8	14	100
TRGVAKAV	1191	8	11	79
VRVCEKMA	2620	8	14	100
VRVLEDGV	155	8	13	93
YRGLDVSV	1423	8	14	100
ARHTPVNSW	2853	9	11	79
ARLIVFPDL	2610	9	11	79
ARLVVLATA	1346	9	11	79
ARMILMTHF	2874	9	12	86
ARPDYNPPL	2298	9	11	79
DRSELSPLL	663	9	11	79

HCV B27 Super Motif

Sequence	Position	Peptide No.	No. of Amino Acids	Sequence Frequency	Conservancy (%)
EKMALYDVV	2624		9	12	86
FKQKALGLL	1733		9	12	86
GHRMAWDMM	315		9	13	93
GKSTKVPAA	1240		9	12	86
GRKPARLIV	2608		9	11	79
HRMAWDMMM	316		9	12	86
IKGGRHLIF	1390		9	11	79
KKKCDELAA	1402		9	14	100
LHGLSAFSL	2919		9	11	79
LHGPTPLLY	1623		9	11	79
LHSYSPGEI	2927		9	11	79
LKGSSGGPL	1166		9	12	86
LRKLGVPPL	2942		9	12	86
NHVSPTHYV	1932		9	12	86
NRRPQDVKF	16		9	11	79
PRRGPRLGV	38		9	13	93
RHTPVNSWL	2854		9	12	86
RHVGPGEGA	1909		9	11	79
RKPARLIVF	2607		9	11	79
RRCRASGVL	2730		9	12	86
RRSRNLGKV	114		9	12	86
SKKKCDELA	1401		9	14	100
THYVPESDA	1937		9	12	86
TKVPAAYAA	1243		9	11	79
TRHADVIPV	1139		9	11	79
TRVESENKV	2251		9	12	86
VKFPGGGQI	22		9	13	93
VRVCEKMAL	2620		9	14	100
WRLLAPITA	1028		9	11	79
WRQEMGGNI	2242		9	12	86
YRGLDVSVI	1423		9	14	100
YRRCRASGV	2729		9	13	93
ARALAHGVRV	148		10	14	100
ARAQAPPPSW	1600		10	11	79
ARHTPVNSWL	2853		10	11	79
ARMILMTHFF	2874		10	12	86
CHSKKKCDEL	1399		10	14	100
DRDRSELSPL	861		10	11	79
DRSELSPLLL	663		10	11	79
EKGGRKPARL	2603		10	11	79
FRAAVCTRGV	1185		10	12	86
GHRMAWDMMM	315		10	12	86
GKSTKVPAAY	1240		10	12	86
GRKPARLIVF	2606		10	11	79
KHMWNFISGI	1768		10	13	93

HCV B27 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
KKCDELAAKL	1403	10	12	86
LHQNIVDVQY	697	10	11	79
LKGSSGGPLL	1168	10	12	86
OKALGLLOTA	1735	10	12	86
RHVGPGEGAV	1909	10	11	79
RRGPRLGVRA	39	10	13	93
RRHVGPGEGA	1908	10	11	79
RRRSRNLGKV	113	10	12	86
RRSRNLGKVI	114	10	12	86
SKFGYGAKDV	2552	10	12	86
SKKKCDELAA	1401	10	14	100
THYVPESDAA	1937	10	12	86
TRGVAKAVDF	1191	10	11	79
TRVESENKVV	2251	10	12	86
VKFPGGGQIV	22	10	13	93
VRVCEKMALY	2620	10	14	100
VRVLEDGVNY	155	10	12	86
WRLLAPITAY	1028	10	11	79
YKVLVLNPSV	1254	10	14	100
YRRCRASGVL	2729	10	12	86
AHGVRVLEDGV	152	11	13	93
AKHMWNFISGI	1767	11	12	86
ARALAHGVRVL	148	11	14	100
ARLIVFPDLGV	2810	11	11	79
CHSKKKCDELA	1399	11	14	100
DRDRSELSPLL	661	11	11	79
EKGGRKPARLI	2603	11	11	79
FRAAVCTRGVA	1185	11	11	79
GKSTKVPAAYA	1240	11	12	86
GKVIDTLTCGF	120	11	12	86
HRMAWDMMMNW	316	11	12	86
KKKCDELAAKL	1402	11	12	86
KRNTNRRPODV	12	11	12	86
LHGPTPLLYRL	1623	11	11	79
LHQNIVDVQYL	697	11	11	79
LKPTLHGPTPL	1619	11	11	79
LRRHVGPGEGA	1907	11	11	79
PRRGPRLGVRA	38	11	13	93
PRRRSRNLGKV	112	11	12	86
RRHVGPGEGAV	1908	11	11	79
RRRSRNLGKVI	113	11	12	86
SRGNHVSPTHY	1929	11	12	86
SRNLGKVIDTL	116	11	12	86
THYVPESDAAA	1937	11	12	86
VRVLEDGVNYA	155	11	12	86

HCV B27 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
YKVLVLNPSVA 136	1254	11	14	100

HCV B58 Super Motif

Table XIII

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
AAILRRHV	1804	8	13	93
AALAAYCL	1673	8	12	86
AAQGYKVL	1250	8	11	79
AATLGFGA	1264	8	14	100
AAVCTRGV	1187	8	12	86
ASLMAFTA	1793	8	11	79
ASSSASQL	2204	8	14	100
ATLGFGAY	1265	8	14	100
CSFSIFLL	172	8	14	100
CSGGAYDI	1310	8	12	86
CSSNVSVA	2819	8	14	100
CTCGSSDL	1128	8	11	79
CTRGVAKA	1190	8	11	79
OTAACGDI	994	8	12	86
DTLTCGFA	124	8	12	86
EAALENLV	750	8	11	79
EAMTRYSA	2794	8	14	100
ESDAAARV	1942	8	12	86
ETAGARLV	1342	8	12	86
ETTMRSPV	1207	8	12	86
FADLMGYI	130	8	13	93
FASRGNHV	1927	8	14	100
FSIFLLAL	174	8	14	100
FSYDTRCF	2670	8	11	79
FTEAMTRY	2792	8	14	100
FTPSPVVV	512	8	13	93
GAGVAGAL	1861	8	12	86
GAHWGVLA	350	8	12	86
GALVVGVV	1885	8	11	79
GARLVVLA	1345	8	12	86
GSGKSTKV	1238	8	13	93
GSSDLYLV	1131	8	12	86
GSSGGPLL	1188	8	12	86
GSSYGFQY	2841	8	11	79
GTFPINAY	2083	8	11	79
HSYSPGEI	2928	8	11	79
HTPVNSWL	2855	8	12	86
ISGIQYLA	1774	8	14	100
ITSCSSNV	2816	8	14	100
ITWGADTA	989	8	12	86
KSTKVPAA	1241	8	12	86
LAGYGAGV	1857	8	11	79
LAHGVRVL	151	8	14	100
LAVAVEPV	972	8	11	79
LSAPSLKA	2211	8	11	79

HCV B58 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
LSPGALVV	1892	8	13	93
LSTGLIHL	690	8	12	86
LTCGFADL	126	8	12	86
LTHIDAHF	1570	8	13	93
MSADLEVV	1654	8	11	79
NSWLGNII	2859	8	14	100
NTCVTQTV	1460	8	12	86
NTNGSWHI	416	8	13	93
PAILSPGA	1889	8	13	93
PALSTGLI	688	8	12	86
PTLWARMI	2870	8	11	79
PTPLLYRL	1628	8	14	100
QATVCARA	1595	8	13	93
RARPRWFM	3019	8	14	100
RSELSPLL	664	8	11	79
RSRNLGKV	115	8	12	86
SAFSLHSY	2923	8	11	79
SSASQLSA	2206	8	14	100
STKVPAAY	1242	8	12	86
STLPGNPA	1784	8	14	100
STLPQAVM	2633	8	12	86
STYGKFLA	1299	8	12	86
TAACGDII	995	8	12	86
TAGARLVV	1343	8	12	86
TTMRSPVF	1208	8	12	86
TTSCGNTL	2739	8	11	79
VAGALVAF	1864	8	12	86
VTRHADVI	1138	8	11	79
VTSTWVLV	1681	8	12	86
WAKHMWNF	1766	8	12	86
WAKVLIVM	368	8	14	100
WAQPGYPW	78	8	12	86
YAAQGYKV	1249	8	11	79
YSIEPLDL	2905	8	11	79
YSTYGKFL	1298	8	12	86
YTNVDQDL	1106	8	11	79
AAKLQDCTM	2758	9	16	114
AAQGYKVLV	1250	9	11	79
AARALAHGV	147	9	11	79
AATLGFGAY	1264	9	14	100
AAVCTRGVA	1187	9	11	79
ASQLSAPSL	2208	9	13	93
ATLGFGAYM	1265	9	26	186
ATVCARAQA	1596	9	11	79
CAAILRRHV	1903	9	13	93

HCV B58 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
CAWYELTPA	1630	9	11	79
CSFSIFLLA	172	9	14	100
CSGGAYDII	1310	9	12	86
CTCGSSDLY	1128	9	11	79
CTRGVAKAV	1190	9	11	79
CTWMNSTGF	555	9	11	79
DAGCAWYEL	1527	9	11	79
DTAACGDII	994	9	12	86
DTRCFDSTV	2673	9	13	93
ETAGARLVV	1342	9	12	86
ETTMRSPVF	1207	9	12	86
FSIFLLALL	174	9	14	100
FSLDPTFTI	1469	9	14	100
FTGLTHIDA	1567	9	13	93
GAGVAGALV	1861	9	12	86
GALVAFKIM	1866	9	12	86
GALVAFKVM	1866	9	14	100
GAVQWMNRL	1916	9	14	100
HSKKKCDEL	1400	9	14	100
HTPGCVPCV	222	9	11	79
ITWGADTAA	989	9	12	86
ITYSTYGKF	1296	9	12	86
KALGLLQTA	1736	9	12	86
KSTKVPAAY	1241	9	12	86
LAALAAYCL	1672	9	12	86
LAEQFKQKA	1729	9	12	86
LAGLAYYSM	356	9	14	100
LAGYGAGVA	1857	9	11	79
LSAFSLHSY	2922	9	11	79
LSTLPGNPA	1783	9	14	100
LTCGFADLM	126	9	24	171
LTDPSHITA	2180	9	14	100
LTGRDKNQV	1052	9	12	86
LTHIDAHFL	1570	9	13	93
LTTSCGNTL	2738	9	11	79
MAKNEVFCV	2592	9	12	86
MAWDMMMNW	318	9	12	86
NAVAYYRGL	1418	9	13	93
NSLLRHHNM	2481	9	14	100
NSWLGNIIM	2858	9	24	171
NTNRRPQDV	14	9	12	86
PAILSPGAL	1889	9	13	93
PSVAATLGF	1261	9	14	100
PTLHGPTPL	1621	9	11	79
PTLWARMIL	2870	9	11	79

HCV B58 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
QAETAGARL	1340	9	12	86
RAAVCTRGV	1186	9	12	86
RALAHGVRV	149	9	14	100
RAQAPPPSW	1601	9	11	79
RAYAMDREM	811	9	16	114
RSELSPLLL	664	9	11	79
RSRNLGKVI	115	9	12	86
SSSASQLSA	2205	9	14	100
STKVPAAYA	1242	9	12	86
STLPGNPAI	1784	9	11	79
STWVLVGGV	1663	9	12	86
TAGARLVVL	1343	9	12	86
TSCSSNVSV	2817	9	14	100
TTIMAKNEV	2589	9	11	79
VAATLGFGA	1263	9	14	100
VAGGHYVQM	933	9	14	100
VAYQATVCA	1592	9	12	86
VAYYRGLDV	1420	9	14	100
VSTLPQAVM	2632	9	12	86
VTQTVDFSL	1463	9	12	86
WAKHMWNFI	1768	9	12	86
YAAQGYKVL	1249	9	11	79
YAPTLWARM	2868	9	14	100
YSPGEINRV	2930	9	11	79
YSPGQRVEF	2848	9	11	79
YSTYGKFLA	1298	9	12	86
YTNVDQDLV	1106	9	11	79
AAQGYKVLVL	1250	10	11	79
AATLGFGAYM	1264	10	28	186
ASLRVFTEAM	2787	10	12	86
ASSSASQLSA	2204	10	14	100
ATGNLPGCSF	165	10	13	93
CSFSIFLLAL	172	10	14	100
CTCGSSDLYL	1128	10	11	79
DARVCACLWM	733	10	18	129
DSVIDCNTCV	1454	10	12	86
DTLTCGFADL	124	10	12	86
EANLLWRQEM	2237	10	24	171
ETAGARLVVL	1342	10	12	86
FADLMGY'FL	130	10	11	79
FTEAMTRYSA	2792	10	14	100
GAARALAHGV	146	10	11	79
GADTAACGDI	992	10	12	86
GAGVAGALVA	1861	10	12	86
GALVVGVVCA	1895	10	11	79

HCV B58 Super Motif:

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
GARLVVLATA	1345	10	11	79
GAVQWMNRLI	1916	10	14	100
GSGKSTKVPA	1238	10	12	86
GTVLDQAETA	1335	10	14	100
HSKKKCDELA	1400	10	14	100
IAFASRGNHV	1925	10	14	100
ISGIQYLAGL	1774	10	14	100
ITRVESENKV	2250	10	12	86
ITSCSSNVSV	2816	10	14	100
ITYSTYGKFL	1296	10	11	79
KSTKVPAAYA	1241	10	12	86
LADGGCSGGA	1305	10	11	79
LAEQFKQKAL	1729	10	12	88
LALPPRAYAM	806	10	12	86
LSPGALVVGV	1892	10	13	93
LSPRGSRPSW	88	10	11	79
LSRARPRWFM	3017	10	14	100
LSTLPGNPAI	1783	10	11	70
LTHPITKYIM	1842	10	16	114
NTCVTQTVDF	1460	10	12	86
PAILSPGALV	1889	10	12	86
PALSTGLIHL	888	10	12	86
PARLIVFPDL	2609	10	11	79
PSWDQMWKCL	1607	10	11	79
PTGSGKSTKV	1236	10	13	93
PTHYVPESDA	1936	10	12	86
PTLHGPTPLL	1621	10	11	79
PTLWARMILM	2870	10	22	157
PTPLLYRLGA	1628	10	13	93
QAETAGARLV	1340	10	12	86
QAPPPSWDQM	1603	10	24	171
QATVCARAQA	1595	10	11	79
RAAKLQDCTM	2757	10	16	114
RAAVCTRGVA	1188	10	11	79
RALAHGVRVL	149	10	14	100
SASQLSAPSL	2207	10	13	93
STKVPAAYAA	1242	10	11	79
STWVLVGGVL	1663	10	12	86
TAGARLVVLA	1343	10	12	86
TARHTPVNSW	2852	10	11	79
TSCSSNVSVA	2817	10	14	100
TSMLTDPSHI	2177	10	13	93
TSTWVLVGGV	1662	10	12	86
TTIMAKNEVF	2589	10	11	79
TTLPALSTGL	685	10	11	79

HCV B58 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
VAATLGFGAY	1263	10	14	100
VTPGERPSGM	1507	10	16	114
VTRHADVIPV	1138	10	11	79
WAQPGYPWPL	76	10	12	86
WARMILMTHF	2873	10	12	86
WARPDYNPPL	2297	10	11	79
YAAQGYKVLV	1249	10	11	79
YSPGEINRVA	2930	10	11	79
YSPGQRVEFL	2648	10	11	79
AARALAHGVRV	147	11	11	79
AASLRVFTEAM	2788	11	12	86
AAVCTRGVAKA	1187	11	11	79
ASHLPYIEQGM	1717	11	14	100
ASQLSAPSLKA	2208	11	11	79
CARAQAPPPSW	1599	11	11	79
CSFSIFLLALL	172	11	14	100
CTCGSSDLYLV	1128	11	11	79
CTRGVAKAVDF	1190	11	11	79
DARVCACLWMM	733	11	16	114
DTLTCGFADLM	124	11	24	171
ETAGARLVVLA	1342	11	12	86
FADLMGYIPLV	130	11	11	79
FSLHSYSPGEI	2925	11	11	79
FTGLTHIDAHF	1567	11	13	93
FTTLPALSTGL	684	11	11	79
GADTAACGDII	992	11	12	86
GAGVAGALVAF	1861	11	12	86
GALVVGVVCAA	1895	11	11	79
GAVQWMNRLIA	1916	11	14	100
GSGKSTKVPAA	1238	11	12	86
HSKKKCDELAA	1400	11	14	100
HSYSPGEINRV	2928	11	11	79
HTPVNSWLGNI	2855	11	12	86
ITRVESENKVV	2250	11	12	86
ITSCSSNVSVA	2816	11	14	100
ITYSTYGKFLA	1296	11	11	79
KSTKVPAAYAA	1241	11	11	79
LADGGCSGGAY	1305	11	11	79
LAGYGAGVAGA	1857	11	11	79
LSNSLLRHHNM	2479	11	14	100
LSPGALVVGVV	1892	11	11	79
LTCGFADLMGY	126	11	12	86
LTSMLTDPSHI	2178	11	13	93
NAVAYYRGLDV	1418	11	13	93
NTNRRPQDVKF	14	11	11	79

<u>HCV B58 Super Motif</u>

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
PAILSPGALVV	1889	11	12	86
PSVAATLGFGA	1261	11	14	100
PTDPRRRSRNL	109	11	12	86
PTHYVPESDAA	1936	11	12	86
PTLHGPTPLLY	1821	11	11	79
PTPLLYRLGAV	1826	11	13	93
QAETAGARLVV	1340	11	12	86
QAPPPSWDQMW	1603	11	11	79
QTVDFSLDPTF	1465	11	12	86
RSQPRGRRQPI	55	11	13	93
SADLEVVTSTW	1655	11	11	79
SSASQLSAPSL	2206	11	13	93
SSDLYLVTRHA	1132	11	12	86
STWVLVGGVLA	1663	11	12	86
TARHTPVNSWL	2852	11	11	79
TSLTGRDKNQV	1050	11	12	86
TSTWVLVGGVL	1662	11	12	86
TTLPALSTGLI	685	11	11	79
VAATLGFGAYM	1263	11	26	106
VAGALVAFKVM	1864	11	14	100
VAVEPVVFSDM	974	11	12	86
VAYQATVCARA	1592	11	11	79
VAYYRGLDVSV	1420	11	14	100
VTSTWVLVGGV	1661	11	12	86
WAQPGYPWPLY	76	11	12	86
WARMILMTHFF	2873	11	12	86
YAAQGYKVLVL	1249	11	11	79
YATGNLPGCSF	164	11	12	86
YTNVDQDLVGW	1106	11	11	79
299				

HCV B62 Super Motif Table XIV

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
AILSPGAL	1880	8	13	93
ALAHGVRV	150	8	14	100
ALGLLQTA	1737	8	12	86
APTLWARM	2869	8	11	79
AQAPPPSW	1602	8	12	86
AQGYKVLV	1251	8	11	79
AVAYYRGL	1419	8	14	100
AVCTRGVA	1180	8	11	79
AVQWMNRL	1817	8	14	100
CLWMMLLI	739	8	12	86
CMSADLEV	1853	8	11	79
CQDHLEFW	1556	8	12	86
CVTQTVDF	1462	8	12	86
DILAGYGA	1855	8	12	86
DLQGSVRL	279	8	12	86
DLMGYIPL	132	8	11	79
DLVNLLPA	1883	8	11	79
DQAETAGA	1339	8	12	86
EIPFYGKA	1377	8	13	93
EQFKQKAL	1731	8	12	86
EVVTSTWV	1659	8	12	86
FISGIQYL	1773	8	14	100
FPDLGVRV	2615	8	11	79
FPGGGQIV	24	8	14	100
FQVAHLHA	1228	8	12	86
GIQYLAGL	1778	8	14	100
GLRDLAVA	960	8	11	79
GPRLGVRA	41	8	13	93
GQIVGGVY	28	8	14	100
GVAGALVA	1863	8	12	86
GVAKAVDF	1193	8	11	79
GVLAALAA	1670	8	12	86
GVRVCEKM	2619	8	14	100
GVVCAAIL	1800	8	11	79
HVGPGEGA	1910	8	11	79
HVSPTHYV	1933	8	12	86
ILGGWVAA	1816	8	12	86
ILGIGTVL	1331	8	12	86
ILSPGALV	1891	8	13	93
IMAKNEVF	2591	8	12	86
IPFYGKAI	1378	8	13	93
IPLVGAPL	137	8	11	79
IVDVQYLY	701	8	12	86
IVFPDLGV	2613	8	11	79
IVGGVYLL	30	8	13	93

HCV B62 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
KMALYDVV	2625	8	12	86
KPARLIVF	2608	8	12	86
KOKALGLL	1734	8	12	86
KVPAAYAA	1244	8	11	79
LIEANLLW	2235	8	12	86
LINTNGSW	414	8	11	79
LLALLSCL	178	8	12	86
LLAPITAY	1030	8	14	100
LLLADARV	729	8	13	93
LLYRLGAV	1629	8	13	93
LMGYIPLV	133	8	11	79
LPALSTGL	687	8	14	100
LPGCSFSI	169	8	13	93
LPRRGPRL	37	8	13	93
LPVCQDHL	1553	8	13	93
LPYIEQGM	1720	8	12	86
LQDCTMLV	2761	8	12	86
LVAYQATV	1591	8	12	86
LVDILAGY	1853	8	11	79
LVGGVLAA	1667	8	12	86
LVLNPSVA	1257	8	14	100
LVNLLPAI	1884	8	11	79
LVTRHADV	1137	8	12	86
LVVGVVCA	1897	8	11	79
LVVICESA	2773	8	11	79
MILMTHFF	2878	8	12	86
MLTDPSHI	2179	8	14	100
NILGGWVA	1815	8	12	86
NIVDVQYL	700	8	12	86
NLLWRQEM	2239	8	12	86
NPSVAATL	1260	8	14	100
PLGGAARA	143	8	11	79
PLLYRLGA	1628	8	13	93
PPPSWDQM	1605	8	12	86
PPSWDQMW	1606	8	11	79
PVVHGCPL	2318	8	11	79
QIVGGVYL	29	8	13	93
QLLRIPQA	336	8	12	86
QPEYDLEL	2808	8	11	79
QPGYPWPL	78	8	12	86
RLHGLSAF	2918	8	12	86
RLIVFPDL	2611	8	11	79
RLLAPITA	1029	8	12	86
RLVVLATA	1347	8	12	86
RMAWDMMM	317	8	12	86

HCV B62 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
RMILMTHF	2875	8	12	86
RPDYNPPL	2299	8	11	79
RQEMGGNI	2243	8	12	86
RVCEKMAL	2621	8	14	100
RVESENKV	2252	8	12	86
RVGDRHYV	2100	8	11	79
SIFLLALL	175	8	14	100
SLDPTFTI	1470	8	14	100
SPGENRV	2931	8	11	79
SPGQRVEF	2649	8	11	79
SQLSAPSL	2209	8	13	93
SVAATLGF	1262	8	14	100
TIMAKNEV	2590	8	11	79
TLGFGAYM	1266	8	13	93
TLHGPTPL	1622	8	11	79
TLPGNPAI	1785	8	11	79
TLWARMIL	2871	8	11	79
TPCSGSWL	1876	8	12	86
TPGCVPCV	223	8	12	86
TQTVDFSL	1464	8	12	86
TVCARAQA	1597	8	11	79
VIDCNTCV	1456	8	12	86
VLAALAAY	1871	8	12	86
VLCECYDA	1521	8	13	93
VLDQAETA	1337	8	14	100
VLEDGVNY	157	8	12	86
VLNPSVAA	1258	8	14	100
VLVGGVLA	1666	8	12	86
VLVLNPSV	1256	8	14	100
VMGSSYGF	2639	8	11	79
VPESDAAA	1940	8	12	86
VQWMNRLI	1918	8	14	100
VVATDALM	1439	8	11	79
VVGVVCAA	1898	8	11	79
VVTSTWVL	1860	8	12	86
WMNRLIAF	1920	8	14	100
WPLLLLLL	799	8	12	86
WVLVGGVL	1665	8	12	86
YLAGLSTL	1779	8	14	100
YPYRLWHY	616	8	14	100
YVPESDAA	1939	8	12	86
AILSPGALV	1890	9	12	86
ALAHGVRVL	150	9	14	100
ALSTGLIHL	689	9	12	86
ALVVGVVCA	1898	9	11	79

HCV B62 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
APPPSWDQM	1604	9	12	86
APTLWARMI	2869	9	11	79
AQGYKVLVL	1251	9	11	79
AQPGYPWPL	77	9	12	86
AVQWMNRLI	1917	9	14	100
CMSADLEVV	1653	9	11	79
DLCGSVFLV	279	9	11	79
DLEVVTSTW	1657	9	12	86
DLMGYIPLV	132	9	11	79
DLVNLLPAI	1883	9	11	79
DLVVICESA	2772	9	11	79
DLYLVTRHA	1134	9	12	86
DPDLSDGSW	2410	9	11	79
DPRRRSRNL	111	9	12	88
EIPFYGKAI	1377	9	13	93
EMGGNITRV	2245	9	12	86
EVVTSTWVL	1658	9	12	86
FISGIQYLA	1773	9	14	100
FLLALLSCL	177	9	12	86
FLLLADARV	728	9	13	93
FQYSPGQRV	2646	9	11	79
GIGTVLDQA	1333	9	14	100
GLPVCQDHL	1552	9	13	93
GLRDLAVAV	968	9	11	79
GLTHIDAHF	1569	9	13	93
GPGEGAVQW	1912	9	12	86
GPTPLLYRL	1625	9	14	100
GQIVGGVYL	28	9	13	93
GVAGALVAF	1863	9	12	86
GVLAALAAY	1670	9	12	86
GVNYATGNL	161	9	11	79
GVRVCEKMA	2619	9	14	100
GVRVLEDGV	154	9	13	93
HLHQNIVDV	696	9	12	86
HLPYIEQGM	1710	9	11	79
HMWNFISGI	1769	9	13	93
HQNIVDVQY	698	9	11	79
HVGPGEGAV	1910	9	11	79
ILAGYGAGV	1656	9	11	79
ILSPGALVV	1891	9	13	93
KVLVLNPSV	1255	9	14	100
LITSCSSNV	2815	9	14	100
LIVFPDLGV	2812	9	11	79
LLFLLLADA	726	9	14	100
LLFNILGGW	1812	9	12	86

HCV B62 Super Motif (No binding data)

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
LLPRRGPRL	38	9	13	93
LPAILSPGA	1888	9	13	93
LPALSTGLI	687	9	12	86
LPCEPEPDV	2165	9	12	86
LPGCSFSIF	169	9	13	93
LVQGVLAAL	1887	9	12	86
LVLNPSVAA	1257	9	14	100
LVNLLPAIL	1894	9	11	79
LVTRHADVI	1137	9	11	79
LVVGVVCAA	1297	9	11	79
NILGGWVAA	1915	9	12	86
NIRTGVRTI	1292	9	11	79
NIVDVQYLY	700	9	12	86
NLGKVIDTL	118	9	12	86
NLPGCSFSI	168	9	13	93
NVDQDLVGW	1108	9	11	79
PLGGAARAL	143	9	11	79
PLLYRLGAV	1628	9	13	93
PPPSWDQMW	1605	9	11	79
PPVVHGCPL	2317	9	11	79
PQPEYDLEL	2807	9	11	79
PVCQDHLEF	1554	9	12	86
PVNSWLGNI	2857	9	14	100
QIVGGVYLL	29	9	13	93
QLSAPSLKA	2210	9	11	79
QPEYDLELI	2808	9	11	79
QPGYPWPLY	78	9	12	86
QPRGRRQPI	57	9	13	93
RLLAPITAY	1029	9	12	86
RMILMTHFF	2875	9	12	86
RVCEKMALY	2621	9	14	100
RVESENKVV	2252	9	12	86
RVLEDGVNY	156	9	12	86
SMLTDPSHI	2178	9	14	100
SPGALVVGV	1893	9	13	93
SPGEINRVA	2931	9	11	79
SPGQRVEFL	2649	9	11	79
SPRGSRPSW	99	9	11	79
SVIDCNTCV	1455	9	12	86
TIMAKNEVF	2590	9	11	79
TLHGPTPLL	1822	9	11	79
TLPALSTGL	686	9	11	79
TLTCGFADL	125	9	12	86
TLWARMILM	2871	9	11	79
TPLLYRLGA	1627	9	13	93

HCV B62 Super Motif

Sequence	Position		No. of Amino Acids	Sequence Frequency	Conservancy (%)
TVLDQAETA	1336		9	14	100
VIDTLTCGF	122		9	12	86
VLEDGVNYA	157		9	12	86
VLVDILAGY	1852		9	11	79
VLVGGVLAA	1666	24.0075	9	12	86
VLVLNPSVA	1258	24.0072	9	14	100
VQWMNRLIA	1918		9	14	100
VVGVVCAAI	1898		9	11	79
VVTSTWVLV	1680	1.0823	9	12	86
WMNRLIAFA	1920	24.0073	9	14	100
WVLVGGVLA	1665	40.0075	9	12	86
YIPLVGAPL	136	1.0817	9	11	79
YLVAYQATV	1590	1.0127	9	12	86
YLVTRHADV	1136	1.0119	9	12	86
YQATVCARA	1594		9	13	93
YVGDLCGSV	276	1.0100	9	12	86
YVGGVEHRL	637	1.0107	9	13	93
YVPESDAAA	1939		9	12	86
AILSPGALVV	1890	24.0101	10	12	86
ALVVGVVCAA	1896		10	11	79
APPPSWDQMW	1604	15.0233	10	11	79
APTLWARMIL	2869	15.0247	10	11	79
AQPGYPWPLY	77		10	12	86
AVAYYRGLDV	1419	1.0486	10	14	100
AVCTRGVAKA	1188		10	11	79
AVQWMNRLIA	1917		10	14	100
CLRKLGVPPL	2941	1.0510	10	12	86
CVTQTVDFSL	1462	1.0487	10	12	86
DILAGYGAGV	1855	1.0495	10	11	79
DLEVVTSTWV	1657	1.0490	10	12	86
DLGVRVCEKM	2617		10	13	93
DLSDGSWSTV	2412	1.0499	10	11	79
DLVNLLPAIL	1883	1.0891	10	11	79
DQAETAGARL	1339		10	12	86
DVKFPGGGQI	21	1174.01	10	12	86
ELITSCSSNV	2814	1.0506	10	14	100
EQRKQKALGL	1731		10	12	86
EVVTSTWVLV	1650	1.0491	10	12	86
GLSAFSLHSY	2921	1.0509	10	11	79
GLSTLPGNPA	1782		10	14	100
GLTHIDAHFL	1569	1.0488	10	13	93
GPGEGAVQWM	1912	15.0240	10	12	86
GQIVGGVYLL	29		10	13	93
GVCWTVYHGA	1081		10	11	79
GVRVCEKMAL	2619	1.0504	10	14	100

HCV B62 Super Motif

Sequence	Position	Peptide No.	No. of Amino Acids	Sequence Frequency	Conservancy (%)
HQNIVDVQYL	698		10	11	79
ILAGYGAGVA	1856		10	11	79
ILGGWVAAQL	1816		10	12	86
IMAKNEVFCV	2591		10	11	79
IQYLAGLSTL	1777		10	14	100
IVFPDLGVRV	2813		10	11	79
KPTLHGPTPL	1620		10	11	79
KVIDTLTCGF	121		10	12	86
KVLVLNPSVA	1255		10	14	100
LLFNILGGWV	1812		10	12	86
LLPAILSPGA	1887		10	13	93
LMGYIPLVGA	133		10	11	79
LPAILSPGAL	1888		10	13	93
LPGCSFSIFL	169		10	13	93
LPRRGPRLGV	37		10	13	93
LPVCQDHLEF	1553		10	12	86
LVAYQATVCA	1591		10	12	86
LVDILAGYGA	1853		10	11	79
LVGGVLAALA	1667		10	12	86
LVVGVVCAAI	1897		10	11	79
MLTDPSHITA	2178		10	14	100
NLPGCSFSIF	168		10	13	93
NPSVAATLGF	1260		10	14	100
PITYSTYGKF	1285		10	11	79
PLGGAARALA	143		10	11	79
PQPEYDLELI	2807		10	11	79
PVCQDHLEFW	1554		10	12	86
PVNSWLGNII	2857		10	14	100
PVYCFTPSPV	508		10	13	93
QLPCEPEPDV	2164		10	12	86
QPEKGGRKPA	2601		10	11	79
RLHGLSAFSL	2918		10	11	79
RLIVFPDLGV	2811		10	11	79
RMAWDMMMNW	317		10	12	86
RVLEDGVNYA	158		10	12	86
SLHSYSPGEI	2826		10	11	79
SLTGRDKNQV	1051		10	12	86
SPGALVVGVV	1893		10	11	79
SQLSAPSLKA	2209		10	11	79
SQPRGRRQPI	56		10	13	93
SVAATLGFGA	1262		10	14	100
TLHGPTPLLY	1822		10	11	79
TLLFNILGGW	1811		10	12	86
TLPALSTGLI	686		10	11	79
TLTCGFADLM	125		10	12	86

HCV B62 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
TPCTCGSSDL	1126	10	11	79
TPLLYRLGAV	1627	10	13	93
TPVNSWLGNI	2856	10	12	86
TVDFSLDPTF	1466	10	12	86
VIDTLTCGFA	122	10	12	86
VLAALAAYCL	1671	10	12	86
VLDQAETAGA	1337	10	12	86
VLNPSVAATL	1258	10	14	100
VLTTSCGNTL	2737	10	11	79
VLVGGVLAAL	1666	10	12	86
VLVLNPSVAA	1256	10	14	100
VMGSSYGFQY	2639	10	11	79
VPESDAAARV	1940	10	12	86
VQWMNRLIAF	1918	10	14	100
VVGVVCAAIL	1898	10	11	79
WVLVGGVLAA	1665	10	12	86
YLKGSSGGPL	1165	10	12	86
YLLPRRGPRL	35	10	13	93
YLVTRHADVI	1136	10	11	79
YVGDLCGSVF	276	10	12	86
ALVVGVVCAAI	1896	11	11	79
APTGSGKSTKV	1235	11	13	93
APTLWARMILM	2869	11	11	79
AQAPPPSWDQM	1602	11	12	86
AVCTRGVAKAV	1188	11	11	79
AVQWMNRLIAF	1917	11	14	100
DILAGYGAGVA	1855	11	11	79
DLEVVTSTWVL	1657	11	12	86
DLGVRVCEKMA	2617	11	13	93
DLMGYIPLVGA	132	11	11	79
DLYLVTRHADV	1134	11	12	86
DQAETAGARLV	1339	11	12	86
DVKFPGGGQIV	21	11	12	86
EQFKQKALGLL	1731	11	12	86
FISGIQYLAGL	1773	11	14	100
FLADGGCSGGA	1304	11	11	79
FPGGGQIVGGV	24	11	14	100
FQYSPGQRVEF	2646	11	11	79
GIQYLAGLSTL	1778	11	14	100
GLPVCQDHLEF	1552	11	12	86
GLSTLPGNPAI	1782	11	11	79
GPTPLLYRLGA	1625	11	13	93
GPVYCFTPSPV	507	11	13	93
GVLAALAAYCL	1670	11	12	86
GVRVCEKMALY	2619	11	14	100

HCV B62 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
GVRVLEDGVNY	154	11	12	86
HLHQNIVDVQY	696	11	11	79
HMWNFISGIQY	1769	11	13	93
HQNIVDVQYLY	698	11	11	79
HVGPGEGAVQW	1910	11	11	79
ILGGWVAAQLA	1610	11	12	86
ILGIGTVLDQA	1331	11	12	86
ILSPGALVVGV	1891	11	13	93
KPARLIVFPDL	2608	11	11	79
KPTLHGPTPLL	1620	11	11	79
KQKALGLLQTA	1734	11	12	86
KVIDTLTCGFA	121	11	12	86
KVLVLNPSVAA	1255	11	14	100
LIAFASRGNHV	1924	11	14	100
LITSCSSNVSV	2815	11	14	100
LIVFPDLGVRV	2612	11	11	79
LLFLLLADARV	726	11	13	93
LLFNILGGWVA	1812	11	12	86
LLPAILSPGAL	1887	11	13	93
LLPRRGPRLGV	36	11	13	93
LLSPRGSRPSW	87	11	11	79
LLWRQEMGGNI	2240	11	12	86
LPAILSPGALV	1888	11	12	86
LPALSTGLIHL	687	11	12	86
LPGCSFSIFLL	169	11	13	93
LPVCQDHLEFW	1553	11	12	86
LVGGVLAALAA	1667	11	12	86
LVLNPSVAATL	1257	11	14	100
LVTRHADVIPV	1137	11	11	79
LVVGVVCAAIL	1897	11	11	79
NILGGWVAAQL	1815	11	12	86
NITRVESENKV	2249	11	12	86
NLLPAILSPGA	1886	11	13	93
NLPGCSFSIFL	180	11	13	93
PITYSTYGKFL	1295	11	11	79
PLEGEPGDPDL	2403	11	13	93
PMGFSYDTRCF	2687	11	11	79
PPSWDQMWKCL	1606	11	11	79
PVNSWLGNIIM	2857	11	12	86
PVYCFTPSPVV	508	11	13	93
RMYVGGVEHRL	635	11	13	93
RQEMGGNITRV	2243	11	12	86
RVCEKMALYDV	2821	11	12	86
SIFLLALLSCL	175	11	12	86
SMLTDPSHITA	2176	11	14	100

HCV B62 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
SPTHYVPESDA	1935	11	12	86
SQLPCEPEPDV	2163	11	12	86
SVAATLGFGAY	1262	11	14	100
TLGFGAYMSKA	1266	11	12	86
TLLFNILGGWV	1811	11	12	86
TPCTCGSSDLY	1126	11	11	79
TPGLPVCQDHL	1550	11	13	93
TPVNSWLGNII	2858	11	12	86
TVLDQAETAGA	1336	11	12	86
VLCECYDAGCA	1521	11	11	79
VLVDILAGYGA	1852	11	11	79
VLVGGVLAALA	1666	11	12	86
VQPEKGGRKPA	2600	11	11	79
VQWMNRLIAFA	1918	11	14	100
VVCAAILRRHV	1901	11	11	79
WVLVGGVLAAL	1665	11	12	86
YLKGSSGGPLL	1165	11	12	86
YLVAYQATVCA	1590	11	12	86
YQATVCARAQA	1594	11	11	79
YVGDLCGSVFL	276	11	12	86
YVPESDAAARV	1939	11	12	86
428				

Table XV **HCV A01 Motif with Binding Information**

Sequence	Position		No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0101
ASFCGSPY	166	26.0026	8	20	100	
DNSVVLSRKY	737	20.0255	10	18	90	0.0001
FAAPFTOCGY	631	20.0254	10	19	95	0.0680
GFAAPFTOCGY	630		11	19	95	
GRETVLEY	140		8	15	75	
GYSLNFMGY	579	2.0058	9	17	85	
HTLWKAGILY	149	1069.04	10	20	100	0.1100
KOAFTFSPTY	653	20.0256	10	19	95	0.0001
LLDTASALY	30	1069.01	9	17	85	12.0000
LSLDVSAAFY	415	1090.07	10	19	95	0.0150
LTFGRETVLEY	137		11	15	75	
MMWYWGPSLY	360	1039.01	10	17	85	0.0810
MSTTDLEAY	103	2.0126	9	15	75	0.8500
NSVVLSRKY	738	2.0123	9	18	90	0.0005
PLDKGIKPY	124	1147.12	9	20	100	
PLDKGIKPYY	124	1069.03	10	20	100	0.1700
PTTGRTSLY	797	1090.09	9	17	85	0.2100
SASFCGSPY	165		9	20	100	
SLDVSAAFY	416	1069.02	9	19	95	5.2000
STTDLEAY	104		8	15	75	
TTGRTSLY	798	26.0030	8	17	85	
WLSLDVSAAFY	414	26.0551	11	19	95	
WMMWYWGPS	359	1039.06	11	17	85	0.3200
YPALMPLY	640	19.0014	8	19	95	
YSLNFMGY	580	26.0032	8	17	85	
25						

Table XVI **HCV A03 Motif with Binding Information**

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0301
AACNWTRGER	647	10	12	86	0.0003
AARALAHGVR	147	10	11	79	
AATLGFGA	1264	8	14	100	
AATLGFGAY	1264	9	14	100	
AAVCTRGVA	1187	9	11	79	
AAVCTRGVAK	1187	10	11	79	
AAVCTRGVAKA	1187	11	11	79	
ACNWTRGER	648	9	12	86	
ADGGCSGGA	1306	9	11	79	
ADGGCSGGAY	1306	10	11	79	
ADVIPVRR	1142	8	12	86	
ADVIPVRRR	1142	9	11	79	
AFASRGNH	1926	8	14	100	
AGALVAFK	1065	8	12	86	
AGARLVVLA	1344	9	12	86	
AGARLVVLATA	1344	11	11	79	
AGLSTLPGNPA	1781	11	14	100	
AGVAGALVA	1062	9	12	86	
AGVAGALVAF	1062	10	12	86	
AGVAGALVAFK	1062	11	12	86	
AGWLLSPR	94	8	12	86	
AGWLLSPRGSR	94	11	12	86	
AGYGAGVA	1050	8	12	86	
AGYGAGVAGA	1050	10	12	86	
ALGLLQTA	1737	8	12	86	
ALSTGLIH	689	8	12	86	
ALSTGLIHLH	689	10	12	86	0.0003
ALVVGVVCA	1896	9	11	79	
ALVVGVVCAA	1896	10	11	79	
ASLMAFTA	1793	8	11	79	
ASQLSAPSLK	2208	10	11	79	
ASQLSAPSLKA	2208	11	11	79	
ASRGNHVSPTH	1928	11	12	86	
ASSSASQLSA	2204	10	14	100	
ATGNLPGCSF	165	10	13	93	
ATLGFGAY	1265	8	14	100	
ATLGFGAYMSK	1265	11	12	86	
ATRKTSER	48	8	11	79	
ATVCARAQA	1596	9	11	79	
AVCTRGVA	1188	8	11	79	
*AVCTRGVAK	1188	9	11	79	0.0260
AVCTRGVAKA	1188	10	11	79	
AVQWMNRLIA	1917	10	14	100	
AVQWMNRLIAF	1917	11	14	100	
CAAILRRH	1903	8	13	93	

HCV A03 Motif with Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0301
CAWYELTPA	1530	9	11	79	
CGFADLMGY	120	9	13	93	
CGNTLTCY	2742	8	11	79	
CGSSDLYLVTR	1130	11	11	79	
CGYRRCRA	2727	8	14	100	
CLRKLGVPPLR	2941	11	12	86	
CSFSIFLLA	172	9	14	100	
CSSNVSVA	2819	8	14	100	
CSSNVSVAH	2819	9	12	86	
CTCGSSDLY	1128	9	11	79	0.0001
CTRGVAKA	1190	8	11	79	
CTRGVAKAVDF	1190	11	11	79	
CTWMNSTGF	555	9	11	79	
CTWMNSTGFTK	555	11	11	79	0.7000
CVQPEKGGR	2599	9	11	79	0.0008
CVQPEKGGRK	2599	10	11	79	0.0011
CVTQTVDF	1462	8	12	86	
DAHFLSQTK	1574	9	14	100	0.0003
DDLVVICESA	2771	10	11	79	
DFSLDPTF	1460	8	14	100	
DGGCSGGA	1307	8	11	79	
DGGCSGGAY	1307	9	11	79	
DIICDECH	1310	9	12	86	
DILAGYGA	1055	8	12	86	
DILAGYGAGVA	1055	11	11	79	
DLGVRVCEK	2617	9	13	93	0.0003
DLGVRVCEKMA	2617	11	13	93	
DLMGYIPLVGA	132	11	11	79	
DLVNLLPA	1883	8	11	79	
DLVVICESA	2772	9	11	79	
DLYLVTRH	1134	8	12	86	
DLYLVTRHA	1134	9	12	86	0.0003
DTLTCGFA	124	8	12	86	
DVIPVRRR	1143	8	11	79	
EAMTRYSA	2794	8	14	100	
ECYDAGCA	1524	8	11	79	
ECYDAGCAWY	1524	10	11	79	
EDLVNLLPA	1882	9	11	79	
EGAVQWMNR	1915	9	14	100	0.0004
EIPFYGKA	1377	8	13	93	
EMGGNITR	2245	8	12	86	
ETAGARLVVLA	1342	11	12	86	
ETTMRSPVF	1207	9	12	86	
EVFCVQPEK	2596	9	12	86	0.0008
FCVQPEKGGR	2598	10	11	79	

<u>HCV A03 Motif with Binding Information</u>

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0301
FCVQPEKGGRK	2590	11	11	79	
FGAYMSKA	1269	8	12	86	
FGAYMSKAH	1269	9	12	86	
FGCTWMNSTGF	553	11	11	79	
FGYGAKDVR	2554	9	12	86	0.0008
FISGIQYLA	1773	9	14	100	
FLADGGCSGGA	1304	11	11	79	
FLLLADAR	728	8	14	100	
FSYDTRCF	2670	8	11	79	
FTEAMTRY	2792	8	14	100	
FTEAMTRYSA	2792	10	14	100	
FTGLTHIDA	1567	9	13	93	
FTGLTHIDAH	1567	10	13	93	
FTGLTHIDAHF	1567	11	13	93	
GAARALAH	146	8	11	79	
GAARALAHGVR	146	11	11	79	
GAGVAGALVA	1861	10	12	86	
GAGVAGALVAF	1861	11	12	86	
GAHWGVLA	350	8	12	86	
GALVVGVVCA	1895	10	11	79	
GALVVGVVCAA	1895	11	11	79	
GARLVVLA	1345	8	12	86	
GARLVVLATA	1345	10	11	79	
GAVQWMNR	1916	8	14	100	
GAVQWMNRLIA	1916	11	14	100	
GAYMSKAH	1270	8	12	86	
GCAWYELTPA	1529	10	11	79	
GCSFSIFLLA	171	10	14	100	
GCTWMNSTGF	554	10	11	79	
GDDLVVICESA	2770	11	11	79	
GDLCGSVF	278	8	12	86	
GFADLMGY	129	8	13	93	
GFGAYMSK	1268	8	12	86	
GFGAYMSKA	1268	9	12	86	
GFGAYMSKAH	1268	10	12	86	
GFQYSPGQR	2645	9	11	79	
GFSYDTRCF	2669	9	11	79	
GGAARALA	145	8	11	79	
GGAARALAH	145	9	11	79	
GGCSGGAY	1300	8	11	79	
GGGQVGGVY	26	10	14	100	
GGHYVQMA	935	8	11	79	
GGQVGGVY	27	9	14	100	
GGRHLIFCH	1392	9	14	100	0.0003
GGRHLIFCHSK	1392	11	14	100	

HCV A03 Motif with Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0301
GGRKPARLIVF	2005	11	11	79	
GGVLAALA	1669	8	12	86	
GGVLAALAA	1669	9	12	86	
GGVLAALAAY	1669	10	12	86	
GGVYLLPR	32	8	13	93	
GGVYLLPRR	32	9	13	93	0.0003
GGWVAAQLA	1818	9	12	86	
GIGTVLDQA	1333	9	14	100	
GIYLLPNR	3037	8	11	79	
GLPVCQDH	1552	8	13	93	
GLPVCQDHLEF	1552	11	12	86	
GLPVSARR	1004	8	11	79	
GLROLAVA	968	8	11	79	
GLSAFSLH	2921	8	11	79	
GLSAFSLHSY	2921	10	11	79	0.0100
GLSTLPGNPA	1782	10	14	100	
GLTHIDAH	1569	8	13	93	
GLTHIDAHF	1569	9	13	93	
GSGKSTKVPA	1238	10	12	86	
GSGKSTKVPAA	1238	11	12	86	
GSSDLYLVTR	1131	10	12	86	
GSSDLYLVTRH	1131	11	12	86	
GSSYGFQY	2641	8	11	79	
GTFPINAY	2063	8	11	79	
GTVLDQAETA	1335	10	14	100	
GVAGALVA	1063	8	12	86	
GVAGALVAF	1063	9	12	86	
GVAGALVAFK	1063	10	12	86	0.3900
GVAKAVDF	1193	8	11	79	
GVCWTVYH	1081	8	11	79	
GVCWTVYHGA	1081	10	11	79	
GVGIYLLPNR	3035	10	11	79	0.0014
GVLAALAA	1670	8	12	86	
GVLAALAAY	1670	9	12	86	0.0046
GVRATRKTSER	45	11	11	79	
GVRVCEKMA	2619	9	14	100	
GVRVCEKMALY	2619	11	14	100	
GVRVLEDGVNY	154	11	12	86	
GVVCAAILR	1900	9	11	79	
GVVCAAILRR	1900	10	11	79	
GVVCAAILRRH	1900	11	11	79	
GVYLLPRR	33	8	13	93	
GVYLLPRRGPR	33	11	13	93	
HADVIPVR	1141	8	11	79	
HADVIPVRR	1141	9	11	79	

HCV A03 Motif with Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0301
HADVIPVRRR	1141	10	11	79	
HAPTGSGK	1234	8	14	100	
HAPTGSGKSTK	1234	11	13	93	
HGLSAFSLH	2920	9	11	79	
HGLSAFSLHSY	2920	11	11	79	
HGPTPLLY	1624	8	11	79	
HGPTPLLYR	1624	9	11	79	
HIDAHFLSQTK	1572	11	14	100	
HLHAPTGSGK	1232	10	12	86	0.5900
HLHQNIVDVQY	696	11	11	79	
HLIFCHSK	1395	8	14	100	
HLIFCHSKK	1395	9	14	100	0.0260
HLIFCHSKKK	1395	10	14	100	0.0260
HMWNFISGIQY	1769	11	13	93	
HSKKKCDELA	1400	10	14	100	
HSKKKCDELAA	1400	11	14	100	
HSYSPGEINR	2928	10	11	79	
HTPGCVPCVR	222	10	11	79	0.0004
HVGPQEGA	1910	8	11	79	
IAFASRGNH	1925	9	14	100	0.0003
IDAHFLSQTK	1573	10	14	100	
IDTLTCGF	123	8	12	86	
IDTLTCGFA	123	9	12	86	
IFCHSKKK	1397	8	14	100	
IGTVLDQA	1334	8	14	100	
IGTVLDQAETA	1334	11	14	100	
IIICDECH	1317	8	12	86	
ILAGYGAGVA	1056	10	11	79	
ILGGWVAA	1816	8	12	86	
ILGGWVAAQLA	1816	11	12	86	
ILGIGTVLDQA	1331	11	12	86	
IMAKNEVF	2591	8	12	86	
ISGIQYLA	1774	8	14	100	
ITRVESENK	2250	9	12	86	0.0150
ITSCSSNVSVA	2816	11	14	100	
ITWGADTA	909	8	12	86	
ITWGADTAA	909	9	12	86	
ITYSTYGK	1296	8	12	86	
ITYSTYGKF	1296	9	12	86	
ITYSTYGKFLA	1296	11	11	79	
IVDVQYLY	701	8	12	86	
IVFPDLGVR	2613	9	11	79	0.0036
IVGGVYLLPR	30	10	13	93	0.0008
IVGGVYLLPRR	30	11	13	93	
KALGLLQTA	1736	9	12	86	

HCV A03 Motif with Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0301
KCDELAAK	1404	8	12	86	
KFGYGAKDVR	2553	10	12	86	
KGGRHLIF	1391	8	11	79	
KGGRHLIFCH	1391	10	11	79	
KGGRKPAR	2604	8	11	79	
KLGVPPLR	2944	8	12	86	
KSTKVPAA	1241	8	12	86	
KSTKVPAAY	1241	9	12	86	0.0009
KSTKVPAAYA	1241	10	12	86	
KSTKVPAAYAA	1241	11	11	79	
KTKRNTNR	10	8	12	86	
KTKRNTNRR	10	9	12	86	0.0110
KTSERSQPR	51	9	13	93	0.1600
KTSERSQPRGR	51	11	12	86	
KVIDTLTCGF	121	10	12	86	
KVIDTLTCGFA	121	11	12	86	
KVLVLNPSVA	1255	10	14	100	
KVLVLNPSVAA	1255	11	14	100	
KVPAAYAA	1244	8	11	79	
LADGGCSGGA	1305	10	11	79	
LADGGCSGGAY	1305	11	11	79	
LAEQFKQK	1729	8	12	86	
LAEQFKQKA	1729	9	12	86	
LAGYGAGVA	1057	9	11	79	
LAGYGAGVAGA	1057	11	11	79	
LCECYDAGCA	1522	10	11	79	
LDQAETAGA	1338	9	12	86	
LDQAETAGAR	1338	10	12	86	
LFLLLADA	727	8	14	100	
LFLLLADAR	727	9	14	100	
LFNILGGWVA	1813	10	12	86	
LFNILGGWVAA	1813	11	12	86	
LFTFSPRR	290	8	11	79	
LGFGAYMSK	1267	9	12	86	0.0810
LGFGAYMSKA	1267	10	12	86	
LGFGAYMSKAH	1267	11	12	86	
LGGAARALA	144	9	11	79	
LGGAARALAH	144	10	11	79	
LGGWVAAQLA	1017	10	12	86	
LGIGTVLDQA	1332	10	13	93	
LGVRATRK	44	8	12	86	
LGVRVCEK	2618	8	14	100	
LGVRVCEKMA	2618	10	14	100	
LIAFASRGNH	1924	10	14	100	
LIEANLLWR	2235	9	12	86	0.0008

HCV A03 Motif with Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0301
LIFCHSKK	1396	8	14	100	
LIFCHSKKK	1396	9	14	100	0.5400
LINTNGSWH	414	9	11	79	
LIVFPDLGVR	2612	10	11	79	0.0003
LLAPITAY	1030	8	14	100	
LLFLLLADA	726	9	14	100	0.0016
LLFLLLADAR	726	10	14	100	
LLFNILGGWVA	1812	11	12	86	
LLPAILSPGA	1807	10	13	93	0.0003
LLPRRGPR	30	8	13	93	
LLSPRGSR	97	8	12	86	
LMGYIPLVGA	133	10	11	79	
LSAFSLHSY	2922	9	11	79	0.0002
LSAPSLKA	2211	8	11	79	
LSNSLLRH	2479	8	12	86	
LSNSLLRHH	2479	9	12	86	0.0003
LSTGLIHLH	690	9	12	86	
LSTLPGNPA	1783	9	14	100	
LTCGFADLMGY	126	11	12	86	
LTDPSHITA	2180	9	14	100	
LTHIDAHF	1570	8	13	93	
LTSMLTDPSH	2176	10	13	93	
LVAYQATVCA	1591	10	12	86	
LVAYQATVCAR	1591	11	11	79	
LVDILAGY	1053	8	11	79	
LVDILAGYGA	1053	10	11	79	
LVGGVLAA	1667	8	12	86	
LVGGVLAALA	1667	10	12	86	
LVGGVLAALAA	1667	11	12	86	
LVLNPSVA	1257	8	14	100	
LVLNPSVAA	1257	9	14	100	
LVVGVVCA	1897	8	11	79	
LVVGVVCAA	1897	9	11	79	
LVVICESA	2773	8	11	79	
MGFSYDTR	2668	8	11	79	
MGFSYDTRCF	2668	10	11	79	
MGSSYGFQY	2640	9	11	79	
MGYIPLVGA	134	9	11	79	
MILMTHFF	2876	8	12	86	
MLTDPSHITA	2179	10	14	100	
MSTNPKPQR	1	9	11	79	
MSTNPKPQRK	1	10	11	79	
NCGYRRCR	2726	8	11	79	
NCGYRRCRA	2726	9	11	79	
NCSIYPGH	305	8	11	79	

HCV A03 Motif with Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0301
NFISGIQY	1772	8	14	100	
NFISGIQYLA	1772	10	14	100	
NGVCWTVY	1080	8	11	79	
NGVCWTVYH	1080	9	11	79	
NGVCWTVYHGA	1080	11	11	79	
NILGGWVA	1815	8	12	86	
NILGGWVAA	1815	9	12	86	
NITRVESENK	2249	10	12	86	0.0010
NIVDVQYLY	700	9	12	86	0.0005
NLLPAILSPGA	1086	11	13	93	
NLPGCSFSIF	160	10	13	93	
NTCVTQTVDF	1460	10	12	86	
NTNRRPQDVK	14	10	11	79	0.0010
NTNRRPQDVKF	14	11	11	79	
NTPGLPVCQDH	1549	11	13	93	
PAILSPGA	1089	8	13	93	
PALSTGLIH	680	9	12	86	
PALSTGLIHLH	680	11	12	86	
PCSGSWLR	1976	8	11	79	
PCTCGSSDLY	1127	10	11	79	
PDLGVRVCEK	2616	10	13	93	
PGALVVGVVCA	1094	11	11	79	
PGCSFSIF	170	8	14	100	
PGCSFSIFLLA	170	11	14	100	
PGCVPCVR	224	8	12	86	
PGEGAVQWMNR	1913	11	13	93	
PGEINRVA	2932	8	11	79	
PGERPSGMF	1509	9	12	86	
PGGGQIVGGVY	25	11	14	100	
PGLPVCQDH	1551	9	13	93	
PGYPWPLY	79	8	14	100	
PITYSTYGK	1295	9	11	79	
PITYSTYGKF	1295	10	11	79	
PLGGAARA	143	8	11	79	
PLGGAARALA	143	10	11	79	
PLGGAARALAH	143	11	11	79	
PLLYRLGA	1628	8	13	93	
PMGFSYDTR	2667	9	11	79	
PMGFSYDTRCF	2667	11	11	79	
PSPVVVGTTDR	514	11	13	93	
PSVAATLGF	1281	9	14	100	
PSVAATLGFGA	1281	11	14	100	
PSWDQMWK	1607	8	11	79	
PTDCFRKH	507	8	13	93	
PTDPRRRSR	109	9	12	86	0.0008

HCY A03 Motif with Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0301
PTGSGKSTK	1236	9	13	93	0.0002
PTHYVPESDA	1936	10	12	86	
PTHYVPESDAA	1936	11	12	86	
PTLHGPTPLLY	1621	11	11	79	
PTPLLYRLGA	1626	10	13	93	
PVQQDHLEF	1554	9	12	86	
PVVVGTTDR	516	9	13	93	0.0008
QAETAGAR	1340	8	12	86	
QATVCARA	1595	8	13	93	
QATVCARAQA	1595	10	11	79	
QIVGGVYLLPR	29	11	13	93	
QLFTFSPR	209	8	12	86	
QLFTFSPRR	209	9	11	79	0.7500
QLLRIPQA	336	8	12	86	
QLSAPSLK	2210	8	11	79	
QLSAPSLKA	2210	9	11	79	
QTVDFSLDPTF	1465	11	12	86	
RAAVCTRGVA	1186	10	11	79	
RAAVCTRGVAK	1186	11	11	79	
RALAHGVR	149	8	14	100	
RATRKTSER	47	9	11	79	
RGNHVSPTH	1930	9	12	86	0.0003
RGNHVSPTHY	1930	10	12	86	0.0003
RGPRLGVR	40	8	13	93	
RGPRLGVRA	40	9	13	93	
RGPRLGVRATR	40	11	11	79	
RGRRQPIPK	59	9	13	93	0.0120
RGSLLSPR	1154	8	12	86	
RGVAKAVDF	1192	9	11	79	
RLGVRATR	43	8	11	79	
RLGVRATRK	43	9	11	79	0.9400
RLHGLSAF	2918	8	12	86	
RLHGLSAFSLH	2918	11	11	79	
RLIAFASR	1923	8	14	100	
RLIAFASRGNH	1923	11	14	100	
RLIVFPDLGVR	2611	11	11	79	
RLLAPITA	1029	8	12	86	
RLLAPITAY	1029	9	12	86	2.7000
RLVVLATA	1347	8	12	86	
RMILMTHF	2075	8	12	86	
RMILMTHFF	2075	9	12	86	
RMYVGGVEH	635	9	14	100	
RMYVGGVEHR	635	10	14	100	0.7200
RSQPRGRR	55	8	13	93	
RVCEKMALY	2621	9	14	100	0.1800

HCV A03 Motif with Binding Information

Sequence	Position		No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0301
RVLEDGVNY	156	1174.17	9	12	86	0.0120
RVLEDGVNYA	156		10	12	86	
SAFSLHSY	2923		8	11	79	
SASQLSAPSLK	2207		11	11	79	
SCSSNVSVA	2818		9	14	100	
SCSSNVSVAH	2818		10	12	86	
SDLYLVTR	1133		8	12	86	
SDLYLVTRH	1133		9	12	86	
SDLYLVTRHA	1133		10	12	86	
SFSIFLLA	173		8	14	100	
SGKSTKVPA	1239		9	12	86	
SGKSTKVPAA	1239		10	12	86	
SGKSTKVPAAY	1239		11	12	86	
SMLTDPSH	2170		8	14	100	
SMLTDPSHITA	2170		11	14	100	
SSASQLSA	2206		8	14	100	
SSDLYLVTR	1132		9	12	86	0.0003
SSDLYLVTRH	1132		10	12	86	0.0003
SSDLYLVTRHA	1132		11	12	86	
SSNVSVAH	2820		8	12	86	
SSSASQLSA	2205		9	14	100	
STGLIHLH	691		8	12	86	
STKVPAAY	1242		8	12	86	
STKVPAAYA	1242		9	12	86	
STKVPAAYAA	1242		10	11	79	
STLPGNPA	1704		8	14	100	
STNPKPQR	2		8	11	79	
STNPKPQRK	2		9	11	79	
STNPKPQRKTK	2		11	11	79	
STWVLVGGVLA	1663		11	12	86	
STYGKFLA	1299		8	12	86	
SVAATLGF	1262		8	14	100	
SVAATLGFGA	1262		10	14	100	
SVAATLGFGAY	1262		11	14	100	
TAGARLVVLA	1343		10	12	86	
TCGFADLMGY	127		10	13	93	
TCGSSDLY	1129		8	11	79	
TCVTQTVDF	1461		9	12	86	
TDPRRRSR	110		8	12	86	
TDPSHITA	2101		8	14	100	
TGEIPFYGK	1375		9	11	79	
TGEIPFYGKA	1375		10	11	79	
TGLTHIDA	1568		8	13	93	
TGLTHIDAH	1568		9	13	93	0.0003
TGLTHIDAHF	1568		10	13	93	

HCV A03 Motif with Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0301
TGNLPGCSF	166	9	13	93	
TGSGKSTK	1237	8	13	93	
TGSGKSTKVPA	1237	11	12	86	
TIMAKNEVF	2590	9	11	79	
TLGFGAYMSK	1266	10	12	86	0.0810
TLGFGAYMSKA	1266	11	12	86	
TLHGPTPLLY	1622	10	11	79	0.0890
TLHGPTPLLYR	1622	11	11	79	
TLPALSTGLIH	686	11	11	79	
TLWARMILMTH	2871	11	11	79	
TSCSSNVSVA	2817	10	14	100	
TSCSSNVSVAH	2817	11	12	86	
TSERSQPR	52	8	13	93	
TSERSQPRGR	52	10	12	86	0.0003
TSERSQPRGRR	52	11	12	86	
TSLTGRDK	1050	8	12	86	
TSMLTDPSH	2177	9	13	93	0.0003
TTIMAKNEVF	2589	10	11	79	
TTMRSPVF	1208	8	12	86	
TVCARAQA	1597	8	11	79	
TVDFSLDPTF	1466	10	12	86	
TVLDQAETA	1336	9	14	100	
TVLDQAETAGA	1336	11	12	86	
VAATLGFGA	1203	9	14	100	
VAATLGFGAY	1203	10	14	100	
VAGALVAF	1064	8	12	86	
VAGALVAFK	1064	9	12	86	0.2400
VAYQATVCA	1592	9	12	86	
VAYQATVCAR	1592	10	11	79	0.0005
VAYQATVCARA	1592	11	11	79	
VCAAILRR	1902	8	11	79	
VCAAILRRH	1902	9	11	79	
VCEKMALY	2622	8	14	100	
VCGPVYCF	505	8	13	93	
VCQDHLEF	1555	8	12	86	
VCTRGVAK	1109	8	11	79	
VCTRGVAKA	1109	9	11	79	
VCWTVYHGA	1002	9	11	79	
VDFSLDPTF	1467	9	14	100	
VDILAGYGA	1854	9	11	79	
VDYPYRLWH	614	9	13	93	
VDYPYRLWHY	614	10	13	93	
VFCVQPEK	2597	8	12	86	
VFCVQPEKGGR	2597	11	11	79	
VFPDLGVR	2614	8	11	79	

HCV A03 Motif with Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0301
VFTGLTHIDA	1566	10	13	93	
VFTGLTHIDAH	1566	11	13	93	
VGDLCGSVF	277	9	12	86	
VGGVLAALA	1668	9	12	86	
VGGVLAALAA	1668	10	12	86	
VGGVLAALAAY	1668	11	12	86	
VGGVYLLPR	31	9	13	93	0.0003
VGGVYLLPRR	31	10	13	93	
VGIYLLPNR	3036	9	11	79	0.0007
VGVVCAAILR	1899	10	11	79	
VGVVCAAILRR	1899	11	11	79	
VIDTLTCGF	122	9	12	86	
VIDTLTCGFA	122	10	12	86	
VLAALAAY	1671	8	12	86	
VLCECYDA	1521	8	13	93	
VLCECYDAGCA	1521	11	11	79	
VLDQAETA	1337	8	14	100	
VLDQAETAGA	1337	10	12	86	
VLDQAETAGAR	1337	11	12	86	
VLEDGVNY	157	8	12	86	
VLEDGVNYA	157	9	12	86	
VLNPSVAA	1258	8	14	100	
VLTSMLTDPSH	2176	11	13	93	
VLVDILAGY	1052	9	11	79	
VLVDILAGYGA	1052	11	11	79	
VLVGGVLA	1666	8	12	86	
VLVGGVLAA	1666	9	12	86	0.0003
VLVGGVLAALA	1666	11	12	86	
VLVLNPSVA	1256	9	14	100	0.0003
VLVLNPSVAA	1256	10	14	100	
VMGSSYGF	2639	8	11	79	
VMGSSYGFQY	2639	10	11	79	
VTRHADVIPVR	1138	11	11	79	
VVCAAILR	1901	8	11	79	
VVCAAILRR	1901	9	11	79	
VVCAAILRRH	1901	10	11	79	
VVGVVCAA	1898	8	11	79	
VVGVVCAAILR	1898	11	11	79	
VVVGTTDR	517	8	13	93	
WAGWLLSPR	93	9	12	86	
WAKHMWNF	1766	8	12	86	
WAQPGYPWPLY	76	11	12	86	
WARMILMTH	2873	9	12	86	
WARMILMTHF	2873	10	12	86	
WARMILMTHFF	2873	11	12	86	

HCV A03 Motif with Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0301
WGPTDPRR	107	8	12	86	
WGPTDPRRR	107	9	12	86	
WGPTDPRRRSR	107	11	12	86	
WLLSPRGSR	98	9	12	86	0.0008
WMNRLIAF	1920	8	14	100	
WMNRLIAFA	1920	9	14	100	0.0003
WMNRLIAFASR	1920	11	14	100	
WMNSTGFTK	557	9	11	79	0.0530
WVLVGGVLA	1665	9	12	86	
WVLVGGVLAA	1665	10	12	86	
YATGNLPGCSF	164	11	12	86	
YDAGCAWY	1526	8	11	79	
YDIIICDECH	1315	10	12	86	
YGAGVAGA	1060	8	12	86	
YGAGVAGALVA	1060	11	12	86	
YGFQYSPGQR	2644	10	11	79	
YLLPRRGPR	35	9	13	93	0.0054
YLVAYQATVCA	1590	11	12	86	
YSPGEINR	2930	8	11	79	
YSPGEINRVA	2930	10	11	79	
YSPGQRVEF	2648	9	11	79	
YSTYGKFLA	1298	9	12	86	
YVGDLCGSVF	276	10	12	86	
YWGGVEHR	637	8	14	100	
YVPESDAA	1939	8	12	86	
YVPESDAAA	1939	9	12	86	
YVPESDAAAR	1939	10	12	86	0.0003
567		3			

Table XVII **HCV A11 Motif With Binding Information**

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*1101
AACNWTRGER	647	10	12	86	0.0140
AARALAHGVR	147	10	11	79	
AATLGFGAY	1264	9	14	100	
AAVCTRGVAK	1107	10	11	79	
ACNWTRGER	648	9	12	86	
ADGGCSGGAY	1306	10	11	79	
ADVIPVRR	1142	8	12	86	
ADVIPVRRR	1142	9	11	79	
AFASRGNH	1926	8	14	100	
AGALVAFK	1065	8	12	86	
AGVAGALVAFK	1062	11	12	86	
AGWLLSPR	94	8	12	86	
AGWLLSPRGSR	94	11	12	86	
ALSTGLIH	689	8	12	86	
ALSTGLIHLH	689	10	12	86	0.0027
ASQLSAPSLK	2208	10	11	79	
ASRGNHVSPTH	1928	11	12	86	
ATLGFGAY	1265	8	14	100	
ATLGFGAYMSK	1265	11	12	86	
ATRKTSER	48	8	11	79	
AVCTRGVAK	1108	9	11	79	0.0250
CAAILRRH	1903	8	13	93	
CGFADLMGY	128	9	13	93	
CGNTLTCY	2742	8	11	79	
CGSSDLYLVTR	1130	11	11	79	
CLRKLGVPPLR	2941	11	12	86	
CNCSIYPGH	304	9	11	79	
CNWTRGER	649	8	12	86	
CSSNVSVAH	2019	9	12	86	
CTCGSSDLY	1128	9	11	79	0.0063
CTWMNSTGFTK	555	11	11	79	0.7500
CVQPEKGGR	2599	9	11	79	0.0005
CVQPEKGGRK	2599	10	11	79	0.0008
DAHFLSQTK	1574	9	14	100	0.0005
DGGCSGGAY	1307	9	11	79	
DIICDECH	1316	9	12	86	
DLGVRVCEK	2617	9	13	93	0.0002
DLYLVTRH	1134	8	12	86	
DVIPVRRR	1143	8	11	79	
ECYDAGCAWY	1524	10	11	79	
EGAVQWMNR	1915	9	14	100	0.0014
EMGGNITR	2245	8	12	86	
EVFCVQPEK	2596	9	12	86	0.0270
FCVQPEKGGR	2598	10	11	79	
FCVQPEKGGRK	2598	11	11	79	

HCV A11 Motif With Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*1101
FGAYMSKAH	1269	9	12	86	
FGYGAKDVR	2554	9	12	86	0.0005
FLLLADAR	728	8	14	100	
FTEAMTRY	2792	8	14	100	
FTGLTHIDAH	1567	10	13	93	
GAARALAH	146	8	11	79	
GAARALAHGVR	146	11	11	79	
GAVQWMNR	1916	8	14	100	
GAYMSKAH	1270	8	12	86	
GFADLMGY	129	8	13	93	
GFGAYMSK	1268	8	12	86	
GFGAYMSKAH	1268	10	12	86	
GFQYSPGQR	2645	9	11	79	
GGAARALAH	145	9	11	79	
GGCSGGAY	1308	8	11	79	
GGGQIVGGVY	26	10	14	100	
GGQIVGGVY	27	9	14	100	
GGRHLIFCH	1392	9	14	100	0.0001
GGRHLIFCHSK	1392	11	14	100	
GGVLAALAAY	1669	10	12	86	
GGVYLLPR	32	8	13	93	
GGVYLLPRR	32	9	13	93	0.0010
GIYLLPNR	3037	8	11	79	
GLPVCQDH	1552	8	13	93	
GLPVSARR	1004	8	11	79	
GLSAFSLH	2921	8	11	79	
GLSAFSLHSY	2921	10	11	79	0.0005
GLTHIDAH	1569	8	13	93	
GNHVSPTH	1931	8	12	86	
GNHVSPTHY	1931	9	12	86	
GNITRVESENK	2248	11	12	86	
GSSDLYLVTR	1131	10	12	86	
GSSDLYLVTRH	1131	11	12	86	
GSSYGFQY	2641	8	11	79	
GTFPINAY	2063	8	11	79	
GVAGALVAFK	1863	10	12	86	1.4000
GVCWTVYH	1081	8	11	79	
GVGIYLLPNR	3035	10	11	79	0.0140
GVLAALAAY	1670	9	12	86	0.0110
GVRATRKTSER	45	11	11	79	
GVRVCEKMALY	2619	11	14	100	
GVRVLEDGVNY	154	11	12	86	
GVVCAAILR	1900	9	11	79	
GVVCAAILRR	1900	10	11	79	
GVVCAAILRRH	1900	11	11	79	

HCV A11 Motif With Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*1101
GVYLLPRR	33	8	13	93	
GVYLLPRRGPR	33	11	13	93	
HADVIPVR	1141	8	11	79	
HADVIPVRR	1141	9	11	79	
HADVIPVRRR	1141	10	11	79	
HAPTGSGK	1234	8	14	100	
HAPTGSGKSTK	1234	11	13	93	
HGLSAFSLH	2920	9	11	79	
HGLSAFSLHSY	2920	11	11	79	
HGPTPLLY	1624	8	11	79	
HGPTPLLYR	1624	9	11	79	
HIDAHFLSQTK	1572	11	14	100	
HLHAPTGSGK	1232	10	12	86	0.0024
HLHQNIVDVQY	696	11	11	79	
HLIFCHSK	1395	8	14	100	
HLIFCHSKK	1305	9	14	100	0.0006
HLIFCHSKKK	1395	10	14	100	0.0002
HMWNFISGIQY	1769	11	13	93	
HSYSPGEINR	2920	10	11	79	
HTPGCVPCVR	222	10	11	79	0.0012
IAFASRGNH	1925	9	14	100	0.0003
IDAHFLSQTK	1573	10	14	100	
IFCHSKKK	1397	8	14	100	
IICDECH	1317	8	12	86	
INTNGSWH	415	8	11	79	
ITRVESENK	2250	9	12	86	0.0079
ITYSTYGK	1208	8	12	86	
IVDVQYLY	701	8	12	86	
IVFPDLGVR	2813	9	11	79	0.0044
IVGGVYLLPR	30	10	13	93	0.0056
IVGGVYLLPRR	30	11	13	93	
KCDELAAK	1404	8	12	86	
KFGYGAKDVR	2553	10	12	86	
KGGRHLIFCH	1391	10	11	79	
KGGRKPAR	2804	8	11	79	
KLGVPPLR	2944	8	12	86	
KNEVFCVQPEK	2594	11	11	79	
KSTKVPAAY	1241	9	12	86	0.0001
KTKRNTNR	10	8	12	86	
KTKRNTNRR	10	9	12	86	0.0100
KTSERSQPR	51	9	13	93	0.0640
KTSERSQPRGR	51	11	12	86	
LADGGCSGGAY	1305	11	11	79	
LAEQFKQK	1729	8	12	86	
LDQAETAGAR	1338	10	12	86	

HCV A11 Motif With Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*1101
LFLLLADAR	727	9	14	100	
LFTFSPRR	290	8	11	79	
LGFGAYMSK	1267	9	12	86	0.2900
LGFGAYMSKAH	1267	11	12	86	
LGGAARALAH	144	10	11	79	
LGVRATRK	44	8	12	86	
LGVRVCEK	2618	8	14	100	
LIAFASRGNH	1924	10	14	100	
LIEANLLWR	2235	9	12	86	0.0005
LIFCHSKK	1390	8	14	100	
LIFCHSKKK	1390	9	14	100	0.1900
LINTNGSWH	414	9	11	79	
LIVFPDLGVR	2612	10	11	79	0.0001
LLAPITAY	1030	8	14	100	
LLFLLLADAR	726	10	14	100	
LLPRRGPR	36	8	13	93	
LLSPRGSR	97	8	12	86	
LSAFSLHSY	2922	9	11	79	0.0002
LSNSLLRH	2479	8	12	86	
LSNSLLRHH	2479	9	12	86	0.0001
LSTGLIHLH	690	9	12	86	
LTCGFADLMGY	126	11	12	86	
LTSMLTDPSH	2176	10	13	93	
LVAYQATVCAR	1591	11	11	79	
LVDILAGY	1053	8	11	79	
MGFSYDTR	2660	8	11	79	
MGSSYGFQY	2640	9	11	79	
MNRLIAFASR	1921	10	14	100	
MNSTGFTK	550	8	11	79	
MSTNPKPQR	1	9	11	79	
MSTNPKPQRK	1	10	11	79	
NCGYRRCR	2726	8	11	79	
NCSIYPGH	305	8	11	79	
NFISGIQY	1772	8	14	100	
NGVCWTVY	1080	8	11	79	
NGVCWTVYH	1080	9	11	79	
NITRVESENK	2249	10	12	86	0.0062
NIVDVQYLY	700	9	12	86	0.0140
NTNRRPQDVK	14	10	11	79	0.0007
NTPGLPVCQDH	1549	11	13	93	
PALSTGLIH	688	9	12	86	
PALSTGLIHLH	688	11	12	86	
PCSGSWLR	1976	8	11	79	
PCTCGSSDLY	1127	10	11	79	
PDLGVRVCEK	2616	10	13	93	

HCV A11 Motif With Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*1101
PGCVPCVR	224	8	12	86	
PGEGAVQWMN	1913	11	13	93	
PGGGQIVGGVY	25	11	14	100	
PGLPVCQDH	1551	9	13	93	
PGYPWPLY	79	8	14	100	
PITYSTYGK	1295	9	11	79	
PLGGAARALAH	143	11	11	79	
PMGFSYDTR	2667	9	11	79	
PNIRTGVR	1281	8	13	93	
PSPVVVGTTDR	514	11	13	93	
PSWDQMWK	1607	8	11	79	
PTDCFRKH	507	8	13	93	
PTDPRRRSR	109	9	12	86	0.0005
PTGSGKSTK	1236	9	13	93	0.0001
PTLHGPTPLLY	1621	11	11	79	
PVVVGTTDR	516	9	13	93	0.0005
QAETAGAR	1340	8	12	86	
QIVGGVYLLPR	28	11	13	93	
QLFTFSPR	289	8	12	86	
QLFTFSPRR	289	9	11	79	0.0330
QLSAPSLK	2210	8	11	79	
QNIVDVQY	699	8	11	79	
QNIVDVQYLY	699	10	11	79	
RAAVCTRGVAK	1100	11	11	79	
RALAHGVR	149	8	14	100	
RATRKTSER	47	9	11	79	
RGNHVSPTH	1930	9	12	86	0.0001
RGNHVSPTHY	1930	10	12	86	0.0001
RGPRLGVR	40	8	13	93	
RGPRLGVRATR	40	11	11	79	
RGRRQPIPK	59	9	13	93	0.0017
RGSLLSPR	1154	8	12	86	
RLGVRATR	43	8	11	79	
RLGVRATRK	43	9	11	79	0.0290
RLHGLSAFSLH	2918	11	11	79	
RLIAFASR	1923	8	14	100	
RLIAFASRGNH	1923	11	14	100	
RLVFPDLGVR	2611	11	11	79	
RLLAPITAY	1029	9	12	86	0.0270
RMYVGGVEH	635	9	14	100	
RMYVGGVEHR	635	10	14	100	0.0200
RNTNRRPQDVK	13	11	11	79	
RSQPRGRR	55	8	13	93	
RVCEKMALY	2621	9	14	100	0.5000
RVLEDGVNY	150	9	12	86	0.0068

HCV A11 Motif With Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*1101
SAFSLHSY	2923	8	11	79	
SASQLSAPSLK	2207	11	11	79	
SCSSNVSVAH	2818	10	12	86	
SDLYLVTR	1133	8	12	86	
SDLYLVTRH	1133	9	12	86	
SGKSTKVPAAY	1239	11	12	86	
SMLTDPSH	2178	8	14	100	
SNSLLRHH	2480	8	12	86	
SSDLYLVTR	1132	9	12	86	0.0044
SSDLYLVTRH	1132	10	12	86	0.0013
SSNVSVAH	2820	8	12	86	
STGLIHLH	691	8	12	86	
STKVPAAY	1242	8	12	86	
STNPKPQR	2	8	11	79	
STNPKPQRK	2	9	11	79	
STNPKPQRKTK	2	11	11	79	
SVAATLGFGAY	1262	11	14	100	
TCGFADLMGY	127	10	13	93	
TCGSSDLY	1129	8	11	79	
TDPRRRSR	110	8	12	86	
TGEIPFYGK	1375	9	11	79	
TGLTHIDAH	1568	9	13	93	0.0001
TGSGKSTK	1237	8	13	93	
TLGFGAYMSK	1268	10	12	86	0.0610
TLHGPTPLLY	1622	10	11	79	0.0007
TLHGPTPLLYR	1622	11	11	79	
TLPALSTGLIH	888	11	11	79	
TLWARMILMTH	2871	11	11	79	
TNPKPQRK	3	8	11	79	
TNPKPQRKTK	3	10	11	79	
TNPKPQRKTKR	3	11	11	79	
TNRRPQDVK	15	9	11	79	
TSCSSNVSVAH	2817	11	12	86	
TSERSQPR	52	8	13	93	
TSERSQPRGR	52	10	12	86	0.0001
TSERSQPRGRR	52	11	12	86	
TSLTGRDK	1050	8	12	86	
TSMLTDPSH	2177	9	13	93	0.0001
VAATLGFGAY	1263	10	14	100	
VAGALVAFK	1864	9	12	86	0.8900
VAYQATVCAR	1592	10	11	79	0.0038
VCAAILRR	1902	8	11	79	
VCAAILRRH	1902	9	11	79	
VCEKMALY	2622	8	14	100	
VCTRGVAK	1189	8	11	79	

HCV A11 Motif With Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*1101
VDYPYRLWH	614	9	13	93	
VDYPYRLWHY	614	10	13	93	
VFCVQPEK	2597	8	12	86	
VFCVQPEKGGR	2597	11	11	79	
VFPDLGVR	2614	8	11	79	
VFTGLTHIDAH	1566	11	13	93	
VGGVLAALAAY	1668	11	12	86	
VGGVYLLPR	31	9	13	93	0.0019
VGGVYLLPRR	31	10	13	93	
VGIYLLPNR	3036	9	11	79	0.0100
VGVVCAAILR	1099	10	11	79	
VGVVCAAILRR	1099	11	11	79	
VLAALAAY	1671	8	12	86	
VLDQAETAGAR	1337	11	12	86	
VLEDGVNY	157	8	12	86	
VLTSMLTDPSH	2175	11	13	93	
VLVDILAGY	1052	9	11	79	
VMGSSYGFQY	2639	10	11	79	
VTRHADVIPVR	1138	11	11	79	
VVCAAILR	1901	8	11	79	
VVCAAILRR	1901	9	11	79	
VVCAAILRRH	1901	10	11	79	
VVGVVCAAILR	1098	11	11	79	
VVVGTTDR	517	8	13	93	
WAGWLLSPR	93	9	12	86	
WAQPGYPWPL	76	11	12	86	
WARMILMTH	2073	9	12	86	
WGPTDPRR	107	8	12	86	
WGPTDPRRR	107	9	12	86	
WGPTDPRRRSR	107	11	12	86	
WLLSPRGSR	96	9	12	86	0.0005
WMNRLIAFASR	1920	11	14	100	
WMNSTGFTK	557	9	11	79	0.0810
WNFISGIQY	1771	9	14	100	
YDAGCAWY	1526	8	11	79	
YDIIICDECH	1315	10	12	86	
YGFQYSPGQR	2644	10	11	79	
YLLPRRGPR	35	9	13	93	0.0005
YSPGEINR	2930	8	11	79	
YVGGVEHR	637	8	14	100	
YVPESDAAAR	1939	10	12	86	0.0001
311		3			

Table XVIII **HCV A24 Motif With Binding Information**

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*2401
AWDMMMNW	319	8	12	86	
AYAAQGYKVL	1248	10	11	79	0.0009
AYYRGLDVSVI	1421	11	14	100	
CYDAGCAW	1525	8	11	79	
CYDAGCAWYEL	1525	11	11	79	
DFSLDPTF	1468	8	14	100	
DFSLDPTFTI	1468	10	14	100	
FWAKHMWNF	1765	9	12	86	6.9000
FWAKHMWNFI	1765	10	12	86	
GFADLMGYI	129	9	13	93	
GFADLMGYIPL	129	11	11	79	
GFSYDTRCF	2669	9	11	79	
GWRLLAPI	1027	8	11	79	
GYGAGVAGAL	1059	10	12	86	0.0003
GYIPLVGAPL	135	10	11	79	0.0057
GYRRCRASGVL	2720	11	12	86	
HMWNFISGI	1769	9	13	93	
IFLLALLSCL	176	10	12	86	
IMAKNEVF	2591	8	12	86	
KFPGGGQI	23	8	13	93	
LFNILGGW	1813	8	12	86	
LWARMILMTHF	2872	11	12	86	
LWRQEMGGNI	2241	10	12	86	
LYLVTRHADVI	1135	11	11	79	
MWNFISGI	1770	8	14	100	
MWNFISGIQYL	1770	11	14	100	
MYVGGVEHRL	636	10	13	93	0.0270
NFISGIQYL	1772	9	14	100	0.0170
PMGFSYDTRCF	2667	11	11	79	
QFKQKALGL	1732	9	12	86	
QFKQKALGLL	1732	10	12	86	
QWMNRLIAF	1910	9	14	100	
QYLAGLSTL	1778	9	14	100	0.0480
QYSPGQRVEF	2647	10	11	79	0.0180
QYSPGQRVEFL	2647	11	11	79	
RMAWDMMMNW	317	10	12	86	
RMILMTHF	2075	8	12	86	
RMILMTHFF	2075	9	12	86	
RMYVGGVEHRL	635	11	13	93	
SFSIFLLAL	173	9	14	100	
SFSIFLLALL	173	10	14	100	0.0041
SMLTDPSHI	2178	9	14	100	
SWDQMWKCL	1600	9	11	79	
SYLKGSSGGPL	1164	11	12	86	
TWMNSTGF	556	8	11	79	

HCV A24 Motif With Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*2401
TWVLVGGVL	1664	9	12	86	
TYSTYGKF	1297	8	13	93	
TYSTYGKFL	1297	9	12	86	0.0230
VFTGLTHI	1566	8	13	93	
VMGSSYGF	2639	8	11	79	
VYLLPRRGPRL	34	11	13	93	0.0016
WMNRLIAF	1920	8	14	100	
YYRGLDVSVI	1422	10	14	100	
53		2			

Table XIX a **HCV DR-Super Motif**

Core Sequence	Core Freq.	Core Conservancy (%)	Exemplary Sequence	Position in HCV Poly-protein	Exemplary Sequence Frequency	Exemplary Sequence Conservancy (%)
FGAYMSKAH	12	86	TLGFGAYMSKAHGVD	1266	5	36
FGCTWMNST	12	86	GNWFGCTWMNSTGFT	550	11	79
FKQKALGLL	12	86	AEQFKQKALGLLQTA	1730	12	86
FLLALLSCL	12	86	FSIFLLALLSCLTVP	174	6	43
FPDLGVRVC	11	79	LVFPDLGVRVCEKM	2612	11	79
FQVAHLHAP	12	86	PQTFQVAHLHAPTGS	1225	6	43
FRAAVCTRG	12	86	VQFRAAVCTRGVAK	1182	7	50
FSIFLLALL	14	100	GCSFSIFLLALLSCL	171	12	86
FSLDPTFTI	14	100	TVDFSLDPTFTIETT	1466	11	79
FTEAMTRYS	14	100	LTVFTEAMTRYSAPP	2769	7	50
FTPSPVVVG	13	93	VYCFTPSPVVVGTTD	509	13	93
FTTLPALST	11	79	PCSFTTLPALSTGLI	681	9	64
FWAKHMWNF	12	86	LEVFWAKHMWNFISG	1702	3	21
IDAHFLSQT	14	100	LTHIDAHFLSQTKQA	1570	7	50
IDCNTCVTQ	12	86	DSVIDCNTCVTQTVD	1454	12	86
IDTLTCGFA	12	86	GKVIDTLTCGFADLM	120	12	86
IEANLLWRQ	12	86	ADLIEANLLWRQEMG	2233	7	50
IFLLALLSC	14	100	SFSIFLLALLSCLTV	173	6	43
ILGGWVAAQ	12	86	LFNILGGWVAAQLAP	1813	8	57
ILGIGTVLD	12	86	STTILGIGTVLDQAE	1328	8	57
LRRHVGPG	11	79	CAWLRRHVGPGEGA	1903	11	79
ILSPGALVV	13	93	LPAILSPGALVVGVV	1888	11	79
INAYTTGPC	12	86	TFPINAYTTGPCTPS	2064	8	57
IPLVGAPLG	11	79	MGYIPLVGAPLGGAA	134	10	71
ITRVESENK	12	86	GGNITRVESENKVVI	2247	10	71
ITSCSSNVS	14	100	LELITSCSSNVSVAH	2813	11	79
IVFPDLGVR	11	79	AHLVFPDLGVRVCE	2610	11	79
LAALAAYCL	12	86	GGVLAALAAYCLTTG	1669	8	57
LADGGCSGG	11	79	GKFLADGGCSGGAYD	1302	10	71
LAGLSTLPG	14	100	IGYLAGLSTLPGNPA	1777	14	100
LAGYGAGVA	11	79	VDILAGYGAGVAGAL	1854	10	71
LATATPPGS	12	86	LVVLATATPPGSVTV	1348	9	64
LDPTFTIET	12	86	DFSLDPTFTIETTTV	1468	5	36
LDQAETAGA	12	86	GTVLDQAETAGARLV	1335	12	86
LELITSCSS	13	93	EYDLELITSCSSNVS	2810	13	93
LEVVTSTWV	12	86	SADLEVVTSTWVLVG	1655	11	79
LFLLLADAR	14	100	VVLLFLLLADARVCS	724	4	29
LGGWVAAQL	12	86	FNILGGWVAAQLAPP	1814	8	57
LGIGTVLDQ	13	93	TILGIGTVLDQAET	1329	9	64
LGVRATRKT	12	86	GPRLGVRATRKTSER	41	10	71
LGVRVCEKM	14	100	FPDLGVRVCEKMALY	2815	11	79
LHGLSAFSL	11	79	IERLHGLSAFSLHSY	2916	6	43
LHGPTPLLY	11	79	KPTLHGPTPLLYRLG	1620	11	79
LHQNIVDVQ	12	86	LHLHQNIVDVQYLY	694	10	71
LHSYSPGEI	11	79	AFSLHSYSPGEINRV	2924	11	79
LIAFASRGN	14	100	MNRLIAFASRGNHVS	1921	12	86
LIEANLLWR	12	86	DADLIEANLLWRQEM	2232	7	50
LIFCHSKKK	14	100	GRHLIFCHSKKKCDE	1393	14	100
LITSCSSNV	14	100	DLELITSCSSNVSVA	2812	13	93
LLALLSCLT	12	86	SIFLLALLSCLTVPA	176	5	36
LLFLLLADA	14	100	YVVLLFLLLADARVC	723	5	36
LLFNILGGW	12	86	QNILLFNILGGWVAA	1809	4	29
LLLADARVC	13	93	LLFLLLADARVCACL	726	9	64
LLPAILSPG	13	93	LVNLLPAILSPGALV	1884	10	71

HCV DR-Super Motif

Core Sequence	Core Freq.	Core Conservancy (%)	Exemplary Sequence	Position in HCV Poly-protein	Exemplary Sequence Frequency	Exemplary Sequence Conservancy (%)
LMGYIPLVG	11	79	FADLMGYIPLVGAPL	130	11	79
LNPSVAATL	14	100	VLVLNPSVAATLGFG	1256	14	100
LPAILSPGA	13	93	VNLLPAILSPGALVV	1865	11	79
LPALSTGLI	12	86	FTTLPALSTGLIHLH	684	11	79
LPRRGPRLG	13	93	VYLLPRRGPRLGVRA	34	13	93
LRDLAVAVE	11	79	HNGLRDLAVAVEPVV	966	4	29
LRKLGVPPL	12	86	ASCLRKLGVPPLRVW	2939	7	50
LSAFSLHSY	11	79	LHGLSAFSLHSYSPG	2919	11	79
LSAPSLKAT	11	79	ASGLSAPSLKATCTT	2208	7	50
LSNSLLRHH	12	86	INALSNSLLRHHNMV	2476	4	29
LSPGALVVG	13	93	PAILSPGALVVGVVC	1889	11	79
LSPLLLSTT	11	79	RSELSPLLLSTTEWQ	664	7	50
LSPRGSRPS	11	79	GWLLSPRGSRPSWGP	95	11	79
LSTGLIHLH	12	86	LPALSTGLIHLHQNI	687	10	71
LTCGFADLM	12	86	IDTLTCGFADLMGYI	123	12	86
LTHIDAHFL	13	93	FTGLTHIDAHFLSQT	1567	13	93
LTSMLTDPS	13	93	VAVLTSMLTDPSHIT	2173	9	64
LVAYQATVC	12	86	FPYLVAYQATVCARA	1508	9	64
LVDILAGYG	11	79	GKVLVDILAGYGAGV	1850	9	64
LVGGVLAAL	12	86	TWVLVGGVLAALAAY	1664	12	86
LVLNPSVAA	14	100	YKVLVLNPSVAATLG	1254	14	100
LVNLLPAIL	11	79	TEDLVNLLPAILSPG	1861	10	71
LVTRHADVI	11	79	DLYLVTRHADVIPVR	1134	11	79
LVVGVVCAA	11	79	PGALVVGVVCAAILR	1094	11	79
LVVLATATP	12	86	GAILVVLATATPPGS	1345	11	79
LWARMILMT	12	86	APTLWARMILMTHFF	2069	11	79
LWRQEMGGN	12	86	ANLLWRQEMGGNITR	2238	12	86
LYRLGAVQN	11	79	IPLLYRLGAVQNEVT	1027	9	64
MAKNEVFCV	12	86	THMAKNEVFCVQPE	2509	9	64
MAWDMMMNW	12	86	GHRMAWDMMMNWSPT	315	12	86
MGGNITRVE	12	86	RQEMGGNITRVESEN	2243	12	86
MGYIPLVGA	11	79	ADLMGYIPLVGAPLG	131	11	79
MLTDPSHIT	14	100	LTSMLTDPSHITAET	2178	8	57
MNRLIAFAS	14.	100	VQWMNRLIAFASRGN	1918	14	100
MTRYSAPPG	14	100	TEAMTRYSAPPGDPP	2793	10	71
MWNFISGIQ	14	100	AKHMWNFISGIQYLA	1767	12	86
MYVGGVEHR	14	100	KVRMYVGGVEHRLNA	633	5	36
VAGALVAFK	12	86	GAGVAGALVAFKVMS	1861	7	50
VAHLHAPTG	12	86	TFQVAHLHAPTGSGK	1227	6	43
VATDALMTG	12	86	VVVVATDALMTGYTG	1437	6	43
VAYQATVCA	12	86	PYLVAYQATVCARAQ	1589	11	79
VCAAILRRH	11	79	VGVVCAAILRRHVGP	1899	10	71
VCEKMALYD	14	100	GVRVCEKMALYDVVS	2619	11	79
VCQDHLEFW	12	86	GLPVCQDHLEFWESV	1552	6	43
VCTRGVAKA	11	79	RAAVCTRGVAKAVDF	1186	11	79
VFCVQPEKG	12	86	KNEVFCVQPEKGGRK	2594	10	71
VFTDNSSPP	11	79	RSPVFTDNSSPPAVP	1211	10	71
VFTGLTHID	13	93	WESVFTGLTHIDAHF	1563	6	43
VGGVLAALA	12	86	WVLVGGVLAALAAYC	1665	12	86
VGGVYLLPR	13	93	GQIVGGVYLLPRRGP	28	13	93
VGSQLPCEP	12	86	GYLVGSQLPCEPEPD	2158	6	43
VGVVCAAIL	11	79	ALVVGVVCAAILRRH	1898	11	79
VIDCNTCVT	12	86	FDSVIDCNTCVTQTV	1453	12	86
VIDTLTCGF	12	86	LQKVIDTLTCGFADL	119	11	79

HCV DR-Super Motif Binding Data Not Included

Core Sequence	Core Freq.	Core Conservancy (%)	Exemplary Sequence	Position in HCV Poly-protein	Exemplary Sequence Frequency	Exemplary Sequence Conservancy (%)
VLAALAAYC	12	86	VGGVLAALAAYCLTT	1668	8	57
VLATATPPG	13	93	RLVVLATATPPGSVT	1347	9	64
VLEDGVNYA	12	86	GVRVLEDGVNYATGN	154	12	86
VLNPSVAAT	14	100	KVLVLNPSVAATLGF	1255	14	100
VLTSMLTDP	13	93	DVAVLTSMLTDPSHI	2172	9	64
VLTTSCGNT	11	79	ASGVLTTSCGNTLTC	2734	10	71
VLVDILAGY	11	79	LGKVLVDILAGYGAG	1849	10	71
VLVGGVLAA	12	86	STWVLVGGVLAALAA	1663	12	86
VLVLNPSVA	14	100	GYKVLVLNPSVAATL	1253	14	100
VNLLPAILS	12	86	EDLVNLLPAILSPGA	1882	11	79
VPESDAAAR	12	86	THYVPESDAAARVTQ	1937	7	50
VTSTWVLVG	12	86	LEVVTSTWVLVGGVL	1658	12	86
VVATDALMT	11	79	DVVVVATDALMTGYT	1436	6	43
VVCAAILRR	11	79	VVGVVCAAILRRHVG	1898	10	71
VVGVVCAAI	11	79	GALVVGVVCAAILRR	1895	11	79
VVLATATPP	12	86	ARLVVLATATPPGSV	1346	9	64
VYCFTPSPV	13	93	CGPVYCFTPSPVVVG	506	13	93
WAGWLLSPR	12	86	GCGWAGWLLSPRGSR	90	5	36
WARMILMTH	12	86	PTLWARMILMTHFFS	2870	11	79
WGADTAACG	12	86	IITWGADTAACGDII	988	6	43
WGPTDPRRR	12	86	RPSWGPTDPRRRSRN	104	10	71
WMNRLIAFA	14	100	AVQWMNRLIAFASRG	1917	14	100
WRLLAPITA	11	79	SKGWRLLAPITAYAQ	1025	4	29
WTGALITPC	11	79	SYTWTGALITPCAAE	2456	9	64
WYELTPAET	12	86	GCAWYELTPAETTVR	1529	5	36
YATGNLPGC	12	86	GVNYATGNLPGCSFS	161	11	79
YCFTPSPVV	13	93	GPVYCFTPSPVVVGT	507	13	93
YDAGCAWYE	11	79	CECYDAGCAWYELTP	1523	10	71
YDIICDEC	12	86	GGAYDIICDECHST	1312	10	71
YDLELITSC	13	93	QPEYDLELITSCSSN	2808	11	79
YGAGVAGAL	12	86	LAGYGAGVAGALVAF	1857	11	79
YGFQYSPGQ	11	79	GSSYGFQYSPGQRVE	2641	10	71
YGKFLADGG	11	79	YSTYGKFLADGGCSG	1298	10	71
YKVLVLNPS	14	100	AQGYKVLVLNPSVAA	1251	11	79
YLAGLSTLP	14	100	GIQYLAGLSTLPGNP	1776	14	100
YLKGSSGGP	12	86	PVSYLKGSSGGPLLC	1162	6	43
YLTRDPTTP	11	79	RVYYLTRDPTTPLAR	2833	9	64
YQATVCARA	13	93	LVAYQATVCARAQAP	1591	11	79
YRGLDVSVI	14	100	VAYYRGLDVSVIPTS	1420	7	50
YRLGAVQNE	11	79	PLLYRLGAVQNEVTL	1628	9	64
YRRCRASGV	13	93	NQGYRRCRASGVLTT	2726	10	71
YSIEPLDLP	11	79	GACYSIEPLDLPQII	2902	6	43
YSPGEINRV	11	79	LHSYSPGEINRVASC	2927	8	57
YVGDLCGSV	12	86	SAMYVGDLCGSVFLV	273	8	57
VGIYLLPNR	11	79		3036		

154

Table XIXb. <u>HCV DR Super Motif With Binding Data</u>

Core Sequence	Exemplary Sequence	DR1	DR2w2 1	DR2w2 2	DR3	DR4w4	DR4w15	DR5w11	DR5w12	DR6w19	DR6w2	DR7	DR9	DRw53
FGAYMSKAI	TLGFGAYMSKAHGVD													
FGCTWMNST	GNWFGCTWMNSTGFT	0.0360	0.0320	0.0013		0.4200	0.0250	0.0210		0.0001	0.0035	0.0250	0.0270	
FKQKALGLL	AEQFKQKALGLLQTA	0.0490				0.0006						0.0058		
FLLALLSCL	FSIFLLALLSCLTVP													
FPDLGVRVC	LVFPDLGVRVCEKM													
FQVAHLHAP	PQTFQVAHLHAPTGS	0.2400				0.0053						-0.0003		
FRAAVCTRG	VQIFRAAVCTRGVAK													
FSIFLLALL	GCSFSIFLLALLSCL	0.0060				0.0015						0.0030		
FSLDPTFTI	TVDFSLDPTFTIETT	0.0001				0.1600						0.0005		
FTEAMTRYS	LTVFTEAMTRYSAPP													
FTPSPVVVG	VYCFTPSPVVVGTTD	0.0180	-0.0001	-0.0003		0.0920	0.0570	0.0056		-0.0001	0.0035	0.0740	0.1800	
FTTLPALST	PCSFTTLPALSTGL													
FWAKHMWNF	LEVFWAKHMWNFISG													
IDAHFLSQT	LTHIDAHFLSQTKQA													
IDCNTCVTQ	DSVIDCNTCVTQTVD	0.0001				-0.0009						-0.0005		
IDTLTCGFA	GKVIDTLTCGFADLM													
IEANLLWRQ	ADLIEANLLWRQEMG													
IFLLALLSC	SFSIFLLALLSCLTV													
ILGGWVAAQ	LFNILGGWVAAQLAP													
ILGIGTVLD	STTILGIGTVLDQAE													
ILRRHVGPG	CAVLRRHVGPGEGA	0.0034				-0.0003						0.0017		
ILSPGALVV	LPAILSPGALVVGVV													
INAYTTGPC	TFPINAYTTGPCTPS													
IPLVGAPLG	MGYIPLVGAPLGGAA													
ITRVESENK	GGNITRVESENKVVI													
IISCSSNVS	LELIISCSSNVSVAH	0.0245	0.0200	-0.0003		0.0070	0.0350	0.0008		0.0510	-0.0003	0.0350	0.0330	
IVFPDLGVR	ARLIVFPDLGVRVCE	0.0053				0.0017						0.0004		
LAALAAYCL	GGVLAALAAYCLTTG													
LAGGCSGG	GKFLAGGCSGGAYD													
LAGLSTLPG	IGYLAGLSTLPGNPA	3.6000	0.0430	0.0094		3.0000		1.7000		-0.0001		0.0021	0.0550	
LAGYGAGVA	VDILAGYGAGVAGAL													
LATATPPGS	LVVLATATPPGSVTV													
LDPTFTIET	DFSLDPTFTIETTTV													
LDDAETAGA	GTVLDDAETAGARLV	0.0001				0.0170						-0.0005		
LELITSCSS	EYDLELITSCSSNVS													
LEVVTSTWV	SADLEVVTSTWVLVG													
LFLLLADAR	VVLLFLLLADARVCS	0.0240				0.0120						0.0033		
LGGWVAAQL	FNILGGWVAAQLAPP													
LGIGTVLDQ	TTILGIGTVLDQAET													
LGVRATRKT	GPRLGVRATRKTSER													
LGVRVCEKM	FPDLGVRVCEKMALY	0.0001				-0.0003						-0.0002		
LHGLSAFSL	IERLHGLSAFSLHSY													
LHGPTPLLY	KPTLHGPTPLLYRLG	0.0380				0.0010						0.0055		
LHQNIVDVQ	LIHLHQNIVDVQYLY													
LISYSPGEI	AFSLISYSPGEINRV	0.0042				-0.0003						0.0024		
LIAFASRGN	MNRLIAFASRGNHVS	0.0760	1.9000	0.0130	0.0058	0.0079	0.0650	0.4400	0.0210	0.4600	0.4800	0.1100	0.2400	
LIEANLLWR	DADLIEANLLWRQEM	0.0008				-0.0010						0.0025		
LIFCHSKKK	GRHLIFCHSKKKCDE	0.0001				-0.0009						-0.0005		
LITSCSSNV	DLELITSCSSNVSVA													
LLALLSCLT	SIFLLALLSCLTVPA													
LLFLLLADA	YVVLLFLLLADARVC													
LLFNILGGW	QHTLLFNILGGWVAA													
LLLADARVC	LLFLLLADARVCACL													
LLPAILSPG	LVNLLPAILSPGALV													
LMGYIPLVG	FADLMGYIPLVGAPL													

HCV DR Super Motif With Binding Data

Core Sequence	Exemplary Sequence	DR1	DR2w2 1	DR2w2 2	DR3	DR4w4	DR4w15	DR5w11	DR5w12	DR6w19	DR8w2	DR7	DR9	DRw53
LNPSVAATL	VLVLNPSVAATLGFG	1.8000	0.0120	0.0004		2.1000	0.0035	0.0140		0.3100	0.0012	1.5000	3.2000	
LPAILSPGA	VHLLPAILSPGALVV													
LPALSTGLI	FTTLPALSTGLIHLH	4.3000	0.0036	0.0016		0.0071		0.0130		0.0002		0.0400	0.0310	
LPRRGPRLG	VYLLPRRGPRLGVRA	0.0140	0.4000	0.0360		-0.0014		0.0120		0.0001		-0.0003	0.0032	
LRDLAVAVE	HNGLRDLAVAVEPVV													
LRKLGVPPL	ASCLRKLGVPPLRVW	1.0000	0.5000	0.0920	0.0051	0.0000	0.4900	0.0310	1.9000	0.0014	0.0730	0.0290	0.0087	
LSAFSLHSY	LHGLSAFSLHSYSPG	1.6000				0.0095						0.0070		
LSAPSLKAT	ASGLSAPSLKATCTT	0.0150				0.0056						0.0008		
LSNSLLRHH	INALSNSLLRHHNMV													
LSPGALVVG	PAILSPGALVVGVVC													
LSPLLLSTT	RSELSPLLLSTTEWQ													
LSPRGSRPS	GWLLSPRGSRPSWGP													
LSTGLIHLH	LPALSTGLIHLHQN													
LTCGFADLM	IDTLTCGFADLMGY	0.0017				0.0024						-0.0003		
LTHIDAHFL	FTGLTHIDAHFLSQT	0.7600	0.0200	0.1300		0.0005	0.0030	0.0003		0.0002	0.0600	0.1400	0.0056	
LTSMLTDPS	VAVLTSMLTDPSHIT													
LVAYQATVC	FPYLVAYQATVCARA													
LVDILAGYG	GKVLVDILAGYGAGV													
LVGGVLAAL	TWVLVGGVLAALAAY	0.7700	0.0011	-0.0003		0.0015		0.0008		0.0001		0.0570	0.0058	
LVLNPSVAA	YKVLVLNPSVAATLG													
LVNLLPAIL	TEGLVNLLPAILSPG													
LVTRHADVI	DLYLVTRHADVIPVR	0.0081	0.0220	0.0011		0.0016		0.0076		0.0005		0.0810	0.0620	
LVVGVVCAA	PGALVVGVVCAALR													
LVVLATATP	GARLVVLATATPPGS	0.0300	0.0009	0.0004		0.8000		0.0094		0.0004		0.0440	0.0067	
LWARMILMT	APTLWARMILMTHFF													
LWRQEMGGN	AHLLWRQEMGGNITR	0.7000				0.0018						0.0022		
LYRLGAVQN	TPLLYRLGAVQNEVT													
MAKNEVFCV	THMAKNEVFCVQPE	0.0014				0.0036						0.0025		
MAWDMMMNW	QLRMAWDMMMNWSPT	0.0200	0.0015	0.0044		0.1600		0.0079		0.0080		0.0017	0.0230	
MGGNITRVE	RQEMGGNITRVESEN	0.0001				-0.0003						-0.0002		
MGYIPLVGA	ADLMGYIPLVGAPLG	0.0006				0.0060						0.0018		
MLTDPSHIT	LTSMLTDPSHITAET	0.0004				0.0740						-0.0003		
MNRLIAFAS	VGWMNRLIAFASRGN													
MIRYSAPPG	TEAMIRYSAPPGDPP													
MWNFISGIQ	AKLMWNFISGIQYLA	1.5000	0.0150	0.0570		0.0040	0.0600	0.0076		0.0004	0.0180	0.2300	0.2700	
MYVGGVEHR	KVRMYVGGVEHRLNA													
VAGALVAFK	GAGVAGALVAFKVMS													
VAHLHAPTG	TFGVAHLHAPTGSGK													
VATDALMTG	VVVVATDALMTGYTG	0.0048	0.0047	0.0014	1.1000			0.0006		0.0029	0.0029	0.0400		
VAYQATVCA	PYLVAYQATVCARAQ													
VCAAILRRH	VGVVCAAILRRHVGP													
VCEKMALYD	GVRVCEKMALYDVVS	0.0022				0.0012						-0.0002		
VCQDHLEFW	GLPVCQDHLEFWESV				0.0063									
VCTRGVAKA	RAAVCTRGVAKAVDF	0.0100				0.0077						0.0024		
VFCVQPEKG	KNEVFCVQPEKGGRK													
VFTDNSSPP	RSPVFTDNSSPPAVP													
VFTGLTHID	WESVFTGLTHIDAHF	0.0310				0.0068						0.0005		
VGGVLAALA	WVLVGGVLAALAAYC													
VGGVYLLPR	GGVGGVYLLPRRGP													
VGSQLPCEP	GYLVGSQLPCEPEPD													
VGVVCAAIL	ALVVGVVCAAILRRH													
VIDCNTCVT	FDSVIDCNTCVTQTV													
VIDTLTCGF	LGKVIDTLTCGFADL	0.0015				0.0096						0.0079		
VLAALAAYC	VGGVLAALAAYCLTT													
VLATATPPG	RLVVLATATPPGSVT													

HCV DR Super Motif With Binding Data

Core Sequence	Exemplary Sequence	DR1	DR2w2 β1	DR2w2 β2	DR3	DR4w4	DR4w15	DR5w11	DR5w12	DR6w19	DR8w2	DR7	DR9	DRw53
VLEDGVNYA	GVRVLEDGVNYATGN	0.0007				0.0006						-0.0002		
VLNPSVAAT	KVLVLNPSVAATLGF													
VLTSMLTDP	DVAVLTSMLTDPSHI													
VLTTSCGNT	ASGVLTTSCGNTLTC													
VLVDILAGY	LGKVLVDILAGYGAG													
VLVGGVLAA	STWVLVGGVLAALAA													
VLVLNPSVA	GYKVLVLNPSVAATL	1.1000	0.0260	0.0004	0.0980	9.6000	0.0670	0.1400	0.0520	0.6900	0.1700	0.2800	1.4000	
VNLLPAILS	EDLVNLLPAILSPGA	0.3700				0.0110						0.0015		
VPESDAAAR	THYVPESDAAARVTQ													
VTSTWVLVG	LEVVTSTWVLVGGVL	0.0120	0.0078	-0.0003		0.0200		0.0008		0.0046		0.1600	0.0120	
VVATDALMT	DVVVVATDALMTGYT	0.0110	0.0110	-0.0003		0.0180	0.0072	-0.0004		0.0140	-0.0003	0.0910	-0.0025	
VVCAAILRR	VVGVVCAAILRRHVG													
VVGVVCAAI	GALVVGVVCAAILRR	0.0170				0.0067						0.0043		
VVLATATPP	ARLVVLATATPPGSV													
VYCFTPSPV	GGPVYCFTPSPVVVG	0.2700	0.0025	-0.0003		0.2600	0.4000	0.0005		-0.0001	0.0011	0.2700	0.4300	
WAGWLLSPR	GCGWAGWLLSPRGSR													
WAIWMLMII	PILWAIWMLMIIFFS	0.0081				0.0200						0.0190		
WGADTAACG	IITWGADTAACGDII													
WGPTDPRRR	RPSWGPTDPRRRSRN													
WMNRLIAFA	AVQWMNRLIAFASRG	2.2000				0.0035						0.0205		
WRLLAPITA	SKGWRLLAPITAYAQ	14.0000	0.0730	0.6600	-0.0006	2.1000	0.2500	4.2000	0.0290	-0.0001	0.9000	0.0260	0.0630	
WTGALITPC	SYTWTGALITPCAAE	0.0260	0.0007	0.0015		0.0680	0.0220	0.0031		-0.0001	0.0130	0.4900	0.0750	
WYELTPAET	GCAWYELTPAETTVR													
YATGNLPGC	GVNYATGNLPGCSFS	0.0011				0.0130						-0.0003		
YCFTPSPVV	GPVYCFTPSPVVVGT													
YDAGCAWYE	CECYDAGCAWYELTP													
YDIRCDEC	GQAYDIRCDECHST													
YDLELITSC	GPEYDLELITSCSSN	0.0003				0.0004						-0.0002		
YGAGVAGAL	LAGYGAGVAGALVAF	0.0410				-0.0003						0.0008		
YGPGYSPGG	GSSYGPGYSPGGRVE	0.4600	0.0001	0.0300	0.0007	0.1200	0.0510	0.0010	0.0003	0.1800	0.0007	0.1600	1.1000	
YGKFLADGG	YSTYGKFLADGGCSG													
YKVLVLNPS	AQGYKVLVLNPSVAA	0.8400	0.0140	0.0004	0.0045	0.3000	0.1700	0.2700	0.0370	0.5900	0.2800	0.0300	0.2000	
YLAGLSTLP	GIGYLAGLSTLPGNP													
YLKGSSGGP	PVSYLKGSSGGPLLC													
YLTRDPTTP	RVYYLTRDPTTPLAR													
YQATVCARA	LVAYQATVCARAQAP													
YRGLDVSVI	VAYYRGLDVSVIPTS													
YRLGAVQNE	PLLYRLGAVQNEVTL													
YRRCRASGV	NCGYRRCRASGVLTT													
YSIEPLDLP	GACYSIEPLDLPQII													
YSPGEINRV	LHSYSPGEINRVASC													
YVGDLCGSV	SAMYVGDLCGSVFLV				-0.0017									
VGIYLLPNR														

HCV DR Super Motif With Binding Data

Core Sequence	Exemplary Sequence	DR1	DR2w2 1	DR2w2 2	DR3	DR4w4	DR4w15	DR5w11	DR5w12	DR6w19	DR8w2	DR7	DR9
VLEDGVNYA	GVRVLEDGVNYATGN	0.0007				0.0006						-0.0002	
VLNPSVAAT	KVLVLNPSVAATLGF												
VLTSMLTDP	DVAVLTSMLTDPSHI												
VLTTSCGNT	ASGVLTTSCGNTLTC												
VLVDILAGY	LGKVLVDILAGYGAG												
VLVGGVLAA	STWVLVGGVLAALAA												
VLVLNPSVA	GYKVLVLNPSVAATL	1.1000	0.0260	0.0004	0.0900	9.6000	0.0670	0.1400	0.0520	0.6900	0.1700	0.2800	1.4000
VNLLPAILS	EDLVNLLPAILSPGA	0.3700				0.0110						0.0015	
VPESDAAAR	THYVPESDAAARVTQ												
VTSTWVLVG	LEVVTSTWVLVGGVL	0.0120	0.0078	-0.0003		0.0200		0.0008		0.0046		0.1600	0.0120
VVATDALMT	DVVVVATDALMTGYT	0.0110	0.0110	-0.0003		0.0180	0.0072	-0.0004		0.0140	-0.0003	0.0910	-0.0025
VVCAAILRR	VVGVVCAAILRRHVG												
VVGVVCAAI	GALVVGVVCAAILRR	0.0170				0.0067						0.0043	
VVLATATPP	ARLVVLATATPPGSV												
VYCFTPSPV	DGPVYCFTPSPVVVG	0.2700	0.0025	-0.0003		0.2600	0.4000	0.0005		-0.0001	0.0011	0.2700	0.4300
WAGWLLSPR	GGGWAGWLLSPRGSR												
WAIIMLMIII	PILWAIIMLMIIFFS	0.0064				0.0200						0.0190	
WGADTAACG	IITWGADTAACGDII												
WGPTDPRRR	RPSWGPTDPRRRSRN												
WMNRLIAFA	AVQWMNRLIAFASRG	2.2000				0.0035						0.0205	
WRLLAPITA	SKGWRLLAPITAYAQ	14.0000	0.0730	0.8800	-0.0006	2.1000	0.2500	4.2000	0.0290	-0.0001	0.9000	0.0260	0.0630
WTGALITPC	SYTWTGALITPCAAE	0.0260	0.0007	0.0015		0.0680	0.0220	0.0031		-0.0001	0.0130	0.4900	0.0750
WYELTPAET	GCAWYELTPAETTVR												
YATGNLPGC	GVNYATGNLPGCSFS	0.0011				0.0130						-0.0003	
YCFTPSPVV	GPVYCFTPSPVVVGT												
YDAGCAWYE	CECYDAGCAWYELTP												
YDIICDEC	GGAYDIICDECHST												
YDLELITSC	QPEYDLELITSCSSN	0.0003				0.0004						-0.0002	
YGAGVAGAL	LAGYGAGVAGALVAF	0.0410				-0.0003						0.0008	
YGPGYSPGG	GSSYGPGYSPGGRVE	0.4600	0.0001	0.0300	0.0007	0.1200	0.0510	0.0010	0.0003	0.1800	0.0007	0.1600	1.1000
YGKFLADGG	YSTYGKFLADGGCSG												
YKVLVLNPS	AQGYKVLVLNPSVAA	0.8400	0.0140	0.0004	0.0045	0.3000	0.1700	0.2700	0.0370	0.5900	0.2800	0.0300	0.2000
YLAGLSTLP	GIGYLAGLSTLPGNP												
YLKGSSGGP	PVSYLKGSSGGPLLC												
YLTRDPTTP	RVYYLTRDPTTPLAR												
YQATVCARA	LVAYQATVCARAQAP												
YRGLDVSVI	VAYYRGLDVSVIPTS												
YRLGAVQNE	PLLYRLGAVQNEVTL												
YRRCRASGV	NGGYRRCRASGVLTT												
YSIEPLDLP	GACYSIEPLDLPQII												
YSPGEINRV	LHSYSPGEINRVASC												
YVGDLCGSV	SAMYVGDLCGSVFLV				-0.0017								
VGIYLLPNR													

154

Table XXb **HCV DR 3A Motif With Binding Information**

Core Sequence	Exemplary Sequence	DR3	DR1	DR2w2β1	DR2w2β2	DR4w4	DR4w15	DR5w11	DR5w12	DR6w19	DR7	DR8w2	DR9	DRw53
FLADGGCSG	YGKFLADGGCSGGAY													
FSLDPTFTI	TVDFSLDPTFTIETT		0.0001			0.1600					0.0005			
LEGEPGDPD	MPPLEGEPGDPDLSD	-0.0017												
LPCEPEPDV	GSQLPCEPEPDVAVL	-0.0017												
MAWDMMMNW	GHRMAWDMMMNWSPT		0.0200	0.0015	0.0044	0.1600		0.0079		0.0080	0.0017		0.0230	
MLTDPSHIT	LTSMLTDPSHITAET		0.0004			0.0740					-0.0003			
MSADLEVVT	MACMSADLEVVTSTW													
VATDALMTG	VVVVATDALMTGYTG	1.1000	0.0048	0.0047	0.0014			0.0006		0.0029	0.0400	0.0029		
VCQDHLEFW	GLPVCQDHLEFWESV	0.0063												
VFPDLGVRV	RLVFPDLGVRVCEK													
VFTDNSSPP	RSPVFTDNSSPPAVP													
VLCECYDAG	DSSVLCECYDAGCAW	-0.0017												
VLEDGVNYA	GVRVLEDGVNYATGN		0.0007			0.0000					-0.0002			
VLVDILAGY	LGKVLVDILAGYGAG													
VQPEKGGRK	VFCVQPEKGGRKPAR													
YDLELITSC	QPEYDLELITSCSSN		0.0003			0.0004					-0.0002			
YSIEPLDLP	GACYSIEPLDLPQII													
YVGDLCGSV	SAMYVGDLCGSVFLV	-0.0017												
YVPESDAAA	PTHYVPESDAAARVT	0.0220												
19														

Table XXc **HCV 3B Motif**

Core Sequence	Core Freq.	Core Conservancy (%)	Exemplary Sequence	Position In HCV Poly-protein	Exemplary Sequence Frequency	Exemplary Sequence Conservancy (%)
FCHSKKKCD	14	100	HLFCHSKKKCDELA	1395	14	100
FSYDTRCFD	11	79	PMGFSYDTRCFDSTV	2667	11	79
LAEQFKQKA	12	86	GMQLAEQFKQKALGL	1726	8	57
LKPTLHGPT	11	79	LRLKPTLHGPTPLL	1616	10	71
VRATRKTSE	11	79	RLGVRATRKTSERSQ	43	10	71
YLVTRHADV	12	86	SDLYLVTRHADVIPV	1133	11	79
MSTNPKPQR 7	11	79		1		

Table XXd **HCV 3B Motif Binding Data**

Core Sequence	Exemplary Sequence	DR1	DR2w2β1	DR2w2β2	DR3	DR4w4	DR4w15	DR5w11	DR5w12	DR6w19	DR8w2	DR7	DR9	DRw53
FQHSKKKCD	HUFCHSKKKCDELA													
FSYDTRCFD	PMGFSYDTRCFDSTV													
LAEQFKQKA	GMCLAEQFKQKALGL				0.0190									
LKPTLHGPT	URLKPTLHGPTPLL													
VRATRKTSE	RLGVRATRKTSERSQ													
YLVTRHADV	SQLYLVTRHADVIPV				0.0022									
MSTNPKPQR7														

TABLE XXI. Population coverage with combined HLA Supertypes

HLA-SUPERTYPES	PHENOTYPIC FREQUENCY Caucasian	North American Black	Japanese	Chinese	Hispanic	Average
a. Individual Supertypes						
A2	45.8	39.0	42.4	45.9	43.0	43.2
A3	37.5	42.1	45.8	52.7	43.1	44.2
B7	38.6	52.7	48.8	35.5	47.1	44.7
A1	47.1	16.1	21.8	14.7	26.3	25.2
A24	23.9	38.9	58.6	40.1	38.3	40.0
B44	43.0	21.2	42.9	39.1	39.0	37.0
B27	28.4	26.1	13.3	13.9	35.3	23.4
B62	12.6	4.8	36.5	25.4	11.1	18.1
B58	10.0	25.1	1.6	9.0	5.9	10.3
b. Combined Supertypes						
A2, A3, B7	83.0	86.1	87.5	88.4	86.3	86.2
A2, A3, B7, A24, B44, A1	99.5	98.1	100.0	99.5	99.4	99.3
A2, A3, B7, A24, B44, A1, B27, B62, B58	99.9	99.6	100.0	99.8	99.9	99.8

SF 184895 v1

Table XXII

HCV ANALOGS

AA	Sequence	Fixed Nomen.	A1 Motif	A2 Super Motif	A3 Super Motif	A24 Motif	B7 Super Motif	1° Anchor Fixer
9	RVXEKMALY		N	N	Y	N	N	
9	AVXTRGVAK		N	N	Y	N	N	
9	EVFXVQPEK		N	N	Y	N	N	
9	HLIFXHSKK		N	N	Y	N	N	
9	LPGXSFSIF		N	N	N	N	Y	
9	LIFXHSKKK		N	N	Y	N	N	
10	VLAALAAYXL		N	Y	N	N	N	No
10	HLIFXHSKKK		N	N	Y	N	N	
10	AAXNWTRGER		N	N	Y	N	N	
10	YLLPRRGPRV	L2.LV10	N	Y	N	N	N	1
9	FPGCSFSIF		N	N	N	N	Y	
9	LPVCSFSIF		N	N	N	N	Y	
9	LPGCSFSYF		N	N	N	N	Y	
9	LPGCMFSIF		N	N	N	N	Y	
9	LPFCSFSIF		N	N	N	N	Y	
9	LPGCSFSPF		N	N	N	N	Y	
9	LPGCSFSII		N	N	N	N	Y	
9	PPVVHGCPI		N	N	N	N	Y	
10	KPTLHGPTPI		N	N	N	N	Y	
10	APTLWARMII		N	N	N	N	Y	
9	SPRGSRPSI		N	N	N	N	Y	
10	LPRRGPRLGI		N	N	N	N	Y	
9	SPGQRVER		N	N	N	N	Y	
9	LPGCSFSII		N	N	N	N	Y	
9	DPRRRSRNI		N	N	N	N	Y	
10	SPGALVVGVI		N	N	N	N	Y	
10	TPLLYRLGAI		N	N	N	N	Y	
9	TISGVLWQV		N	Y	N	N	N	No
9	SISGVLWQV		N	Y	N	N	N	No
9	SLMAFTASV		N	Y	N	N	N	No
9	GLRDCTMLV		N	Y	N	N	N	No
10	KLVALGVNAV		N	Y	N	N	N	No
10	YLLPSRGPKL		N	Y	N	N	N	No
10	KLSGLGLNAV		N	Y	N	N	N	No
10	YVLPRRGPRL	LV2.L10	N	Y	N	N	N	Rev
10	VFFNILGGWV		N	N	N	N	N	
10	KLVSLGVNAV		N	Y	N	N	N	No
9	CINGVCWTA	I2.VA9	N	Y	N	N	N	Rev
9	CANGVCWTV	IA2.V9	N	Y	N	N	N	Rev

HCV ANALOGS

AA	Sequence	Fixed Nomen.	A1 Motif	A2 Super Motif	A3 Super Motif	A24 Motif	B7 Super Motif	1° Anchor Fixer
9	CVNGVCWAV 40		N	Y	N	N	N	

Table XXIII. Immunogenicity of identified supermotif-bearing peptides

					Immunogenicity						
					Human[a]					Transgenic mice[b]	
Supermotif	Peptide	Sequence	Protein	Position	Barnaba; patients	Barnaba; contacts	Chisari	Pape	overall	Frequency	Response
A2	1073.05	LLFNILGGWV	NS4	1812	1/6	7/17	2/21	0/6	10/50	6/6	6.4 (1.7)
	1090.18	FLLLADARV	NS1/E2	728	2/6	7/17	1/21	0/6	10/50	5/6	9.5 (3.0)
	1013.02	YLVAYQATV	NS4	1590	1/6	4/17	1/21	0/6	6/50	5/6	8.5 (3.7)
	1090.22	RLIVFPDLGV	NS5	2578	2/6	5/17	0/21	0/6	7/50	0/6	-
	1013.1002	DLMGYIPLV	Core	132	2/6	7/17	1/21	1/6	11/50	5/6	8.8 (2.6)
	24.0073	WMNRLIAFA	NS4	1920	1/6	3/17	2/21	1/6	7/50	0/6	-
	24.0075	VLVGGVLAA	NS4	1666	1/6	6/17	3/21	1/6	11/50	0/6	-
	1174.08	HMWNFISGI	NS4	1769	3/6	3/17	2/21	0/6	8/50	6/6	6.4 (1.7)
	1073.06	ILAGYGAGV	NS4	1851	2/6	3/17	0/21	0/6	5/50	3/6	54.7 (3.3)
	1073.07	YLLPRRGPRL	CORE	35	2/6	5/17	7/21	1/6	17/50	4/6	59.1 (7.2)
	24.0071	LLFLLLADA	NS1/E2	726	2/6	9/17	0/21	0/6	11/50	0/6	-
	1.0119	YLVTRHADV	NS3	1131	6/6	10/17	0/21	1/6	17/50	0/6	-
A3	1.0952	KTSERSQPR	CORE	51	2/16	1/4	3/12	0/6	6/38	3/6	23.4 (1.3)
	1073.11	RLGVRATRK	CORE	43	4/16	1/4	7/12	1/6	13/38	3/6	42.2 (1.2)
	1.0955	QLFTFSPRR	ENV	290	1/16	0/4	6/12	1/6	8/38		
	1073.13	RMYVGGVEHR	NS1/E2	632	5/16	1/4	4/12	1/6	11/38	2/6	2.8 (1.1)
	1.0123	LIFCHSKKK	NS3	1396	6/16	1/4	4/12	2/6	13/38	3/6	4.4 (1.1)
	1073.10	GVAGALVAFK	NS4	1863	3/16	0/4	6/12	2/6	11/38	6/6	56.5 (1.7)
	24.0090	VAGALVAFK	NS4	1864	4/16	1/4	6/12	0/4	11/38	1/6	7.1
	24.0086	TLGFGAYMSK	NS3	1262	6/16		2/12	2/5	10/33		
B7	1145.12	LPGCSFSIF	CORE	169			2	3/10	5		

Table XXIV. Human and murine MHC-peptide binding assays established using purified MHC molecules and gel filtration chromatography

A. Class I binding assays

Species	Antigen	Allele	Cell line	Radiolabeled peptide: Source	Radiolabeled peptide: Sequence	Notes
Human	A1	A*0101	Steinlin	Hu. J chain 102-110	YTAVVPLVY	no NEN in PI cocktail
	A2	A*0201	JY	HBVc 18-27 F6->Y	FLPSDYFPSV	"
	A2	A*0202	P815 (transfected)	HBVc 18-27 F6->Y	FLPSDYFPSV	"
	A2	A*0203	FUN	HBVc 18-27 F6->Y	FLPSDYFPSV	"
	A2	A*0206	CLA	HBVc 18-27 F6->Y	FLPSDYFPSV	"
	A2	A*0207	721.221 (transfected)	HBVc 18-27 F6->Y	FLPSDYFPSV	"
	A3		GM3107	non-natural (A3CON1)	KVFPYALINK	"
	A11		BVR	non-natural (A3CON1)	KVFPYALINK	"
	A24	A*2402	KAS116	non-natural (A24CON1)	AYIDNYNKF	"
	A31	A*3101	SPACH	non-natural (A3CON1)	KVFPYALINK	"
	A33	A*3301	LWAGS	non-natural (A3CON1)	KVFPYALINK	"
	A28/68	A*6801	C1R	HBVc 141-151 T7->Y	STLPETYVVRR	"
	A28/68	A*6802	AMAI	HBV pol 646-654 C4->A	FTQAGYPAL	"
	B7	B*0702	GM3107	A2 sigal seq. 5-13 (L7->Y)	APRTLVYLL	"
	B8	B*0801	Steinlin	HVgp 586-593 Y1->F, Q5->	FLKDYQLL	"
	B27	B*2705	LG2	R 60s	FRYNGLIHR	"
	B35	B*3501	C1R, BVR	non-natural (B35CON2)	FPFKYAAAF	"
	B35	B*3502	TISI	non-natural (B35CON2)	FPFKYAAAF	"
	B35	B*3503	EHM	non-natural (B35CON2)	FPFKYAAAF	"
	B44	B*4403	PITOUT	EF-1 G6->Y	AEMGKYSFY	"
	B51		KAS116	non-natural (B35CON2)	FPFKYAAAF	"
	B53	B*5301	AMAI	non-natural (B35CON2)	FPFKYAAAF	"
	B54	B*5401	KT3	non-natural (B35CON2)	FPFKYAAAF	"
	Cw4	Cw*0401	C1R	non-natural (C4CON1)	QYDDAVYKL	"
	Cw6	Cw*0602	721.221 transfected	non-natural (C6CON1)	YRHDGGNVL	"
	Cw7	Cw*0702	721.221 transfected	non-natural (C6CON1)	YRHDGGNVL	"
Mouse	D^b		EL4	Adenovirus E1A P7->Y	SGPSNTYPEI	"
	K^b		EL4	VSV NP 52-59	RGYVFQGL	"
	D^d		P815	HIV-IIIB ENV G4->Y	RGPYRAFVTI	"
	K^d		P815	non-natural (KdCON1)	KFNPMKTYI	"
	L^d		P815	HBVs 28-39	IPQSLDSYWTSL	"

Table XXIV. Human and murine MHC-peptide binding assays established using purified MHC molecules and gel filtration chromatography

B. Class II binding assays

Species	Antigen	Allele	Cell line	Radiolabeled peptide		Notes
				Source	Sequence	
Human	DR1	DRB1*0101	LG2	HA Y307-319	YPKYVKQNTLKLAT	
	DR2	DRB1*1501	L466.1	MBP 88-102Y	VVHFFKNIVTPRTPPY	
	DR2	DRB1*1601	L242.5	non-natural (760.16)	YAAFAAAKTAAAFA	
	DR3	DRB1*0301	MAT	MT 65kD Y3-13	YKTIAFDEEARR	optimal assay pH is 4.5
	DR4w4	DRB1*0401	Preiss	non-natural (717.01)	YARFQSQTTLKQKT	
	DR4w10	DRB1*0402	YAR	non-natural (717.10)	YARFQRQTTLKAAA	
	DR4w14	DRB1*0404	BIN 40	non-natural (717.01)	YARFQSQTTLKQKT	
	DR4w15	DRB1*0405	KT3	non-natural (717.01)	YARFQSQTTLKQKT	
	DR7	DRB1*0701	Pitout	Tet. tox. 830-843	QYIKANSKFIGITE	
	DR8	DRB1*0802	OLL	Tet. tox. 830-843	QYIKANSKFIGITE	
	DR8	DRB1*0803	LUY	Tet. tox. 830-843	QYIKANSKFIGITE	
	DR9	DRB1*0901	HID	Tet. tox. 830-843	QYIKANSKFIGITE	
	DR11	DRB1*1101	Sweig	Tet. tox. 830-843	QYIKANSKFIGITE	
	DR12	DRB1*1201	Herluf	unknown eluted peptide	EALIHQLKINPYVLS	
	DR13	DRB1*1302	H0301	Tet. tox. 830-843 S->A	QYIKANAKFIGITE	
	DR51	DRB5*0101	GM3107 or L416.3	Tet. tox. 830-843	QYIKANAKFIGITE	
	DR51	DRB5*0201	L255.1	HA 307-319	PKYVKQNTLKLAT	
	DR52	DRB3*0101	MAT	Tet. tox. 1272-1284	NGQIGNDPNRDIL	
	DR53	DRB4*0101	L257.6	non-natural (717.01)	YARFQSQTTLKQKT	no NEM in PI mix
	DQ3.1	DQA1*0301/DQB1*0301	PF	non-natural (ROIV)	YAHAAHAAHAAHAAHAA	
Mouse	IA^b		DB27.4	non-natural (ROIV)	YAHAAHAAHAAHAAHAA	optimal assay pH is 5.5
	IA^d		A20	non-natural (ROIV)	YAHAAHAAHAAHAAHAA	
	IA^k		CH-12	HEL 46-61	YNTDGSTDYGILQINSR	optimal assay pH is 5.0
	IA^s		LS102.9	non-natural (ROIV)	YAHAAHAAHAAHAAHAA	
	IA^u		91.7	non-natural (ROIV)	YAHAAHAAHAAHAAHAA	
	IE^d		A20	Lambda repressor 12-26	YLEDARRKKAIYEKKK	optimal assay pH is 5.0
	IE^k		CH-12	Lambda repressor 12-26	YLEDARRKKAIYEKKK	optimal assay pH is 5.0

Table XXV. Monoclonal antibodies used in MHC purification.

Monoclonal antibody	Specificity
W6/32	HLA-class I
B123.2	HLA-B and C
IVD12	HLA-DQ
LB3.1	HLA-DR
M1/42	H-2 class I
28-14-8S	H-2 D^b and L^d
34-5-8S	H-2 D^d
B8-24-3	H-2 K^b
SF1-1.1.1	H-2 K^d
Y-3	H-2 K^b
10.3.6	H-2 IA^k
14.4.4	H-2 IE^d, IE^K
MKD6	H-2 IA^d
Y3JP	H-2 IA^b, IA^s, IA^u

Table XXVI: HCV-derived conserved high algorithm A*0201-binding peptides

					A2-supertype binding capacity (IC50 nM)					
Peptide	Molecule	1st Position	Sequence	Consv.	A*0201	A*0202	A*0203	A*0206	A*6802	A2 XRN
1073.05	NS4	1812	LLFNILGGWV	85	4.2	113	3.2	19	33	5
1090.18	NS1/E2	728	FLLLADARV	92	18	90	149	247	111	5
1013.02	NS4	1590	YLVAYQATV	85	20	39	16	82	33	5
1090.22	NS5	2611	RLIVFPDLGV	79	56	391	10	370	8000	4
1013.1002	CORE	132	DLMGYIPLV	79	80	4778	204	481	12	4
24.0073	NS4	1920	WMNRLIAFA	100	122	130	3.3	1609	400	4
24.0075	NS4	1666	VLVGGVLAA	85	185	331	32	308	3077	4
1174.08	NS4	1769	HMWNFISGI	92	15	10750	77	132	7547	3
1073.06	NS4	1851	ILAGYGAGV	79	116	143	5.0	755	889	3
1073.07	CORE	35	YLLPRRGPRL	92	125	6143	455	416	10256	3
24.0071	NS1/E2	726	LLFLLLADA	100	217	287	455	3364	3077	3
1.0119	LORF	1131	YLVTRHADV	85	455	2048	3.6	71	3077	3
24.0065	NS4	1891	ILSPGALVV	92	238	10750	27	1028	3077	2
1013.12	NS1/E2	686	ALSTGLIHL	85	313	7167	45	18500	10256	2
939.14	NS1/E2	696	HLHQNIVDV	85	500	3071	19	1370	10811	2
1090.21	NS5	2918	RLHGLSAFSL	79	179	782	625	18500	12500	1

Table XXVII: HCV-derived conserved high algorithm A*03 and/or A*11 binding peptides

					A3-supertype binding capacity (IC50 nM)					
Peptide	Molecule	1st Position	Sequence	Consv.	A*03	A*11	A*3101	A*3301	A*6801	A3 XRN
1.0952	CORE	51	KTSERSQPR	92	69	94	67	1813	145	4
1073.11	CORE	43	RLGVRATRK	79	12	207	429	-	-	3
1.0955	ENV1	290	QLFTFSPRR	79	15	182	621	3766	3	3
1073.13	NS1/E2	632	RMYVGGVEHR	100	15	300	95	9667	1778	3
1.0123	NS3	1396	LIFCHSKKK	100	20	32	2535	24167	333	3
1073.10	NS4	1863	GVAGALVAFK	85	28	4	3273	26364	118	3
24.0090	NS4	1864	VAGALVAFK	85	46	7	3750	11600	258	3
24.0086	NS3	1262	LGFGAYMSK	85	136	21	2950	22308	222	3
1174.16	NS1/E2	557	WMNSTGFTK	79	208	74	12857	690	1429	2
1073.14	NS3	1261	TLGFGAYMSK	85	136	98	-	22308	8889	2
1090.23	LORF	1183	AVCTRGVAK	79	423	240	16364	-	-	2
1090.24	NS5	2596	EVFCVQPEK	85	13750	222	-	-	18	2
24.0103	NS1/E2	647	AACNWTRGER	85	36667	429	400	5273	4444	2
1073.16	NS3	1232	HLHAPTGSGK	85	19	2500	-	-	2857	1
1073.12	NS3	1395	HLIFCHSKKK	100	423	-	20000	-	-	1
1090.26	NS3	1395	HLIFCHSKK	100	440	10000	-	-	8000	1

* A dash indicates IC50nM >30,000

Table XXVIII: HCV derived conserved B*0702 binding peptides

A. High conservancy 9- and 10-mer peptides.

					B7-supertype binding capacity (IC50 nM)					
Peptide	Molecule	1st Position	Sequence	Consv.	B*0702	B*3501	B*51	B*5301	B*5401	B7 XRN
1145.12	Core	169	LPGCSFSIF	92	28	90	100	114	6667	4
15.0048	E2	681	LPALSTGLI	85	157	-	2.8	1500	20000	2
15.0234	NS3	1620	KPTLHGPTPL	79	3.9	-	27500	-	-	1
15.0247	NS5	2835	APTLWARMIL	79	6.3	-	5500	-	-	1
15.0042	CORE	99	SPRGSRPSW	79	14	-	11000	-	-	1
15.0039	Core	57	QPRGRRQPI	92	24	-	-	-	-	1
15.0218	Core	37	LPRRGPRLGV	92	29	-	6111	-	4000	1
15.0060	NS5	2615	SPGQRVEFL	79	46	-	27500	-	-	1
15.0043	Core	111	DPRRRSRNL	85	324	-	-	-	-	1
15.0063	NS5	2835	APTLWARMI	79	344	-	4583	-	-	1
1292.17	NS5	2317	PPVVHGCPL	79	393	-	-	-	-	1
15.0239	NS4	1893	SPGALVVGVV	79	423	-	3438	-	-	1
15.0235	NS3	1621	TPLLYRLGAV	92	458	-	6875	-	909	1

Table XXVIII: HCV derived conserved B*0702 binding peptides

B. Additional HCV derived B7 supermotif peptides.

					B7-supertype binding capacity (IC50 nM)					
Peptide	Molecule	1st Position	Sequence	Consv.	B*0702	B*3501	B*51	B*5301	B*5401	B7 XRN
29.0035	NS3	1378	IPFYGKAI	92	458	-	46	-	50	3
29.0040	Core	37	LPRRGPRL	92	0.85	-	306	-	5000	2
29.0036	Core	137	IPLVGAPL	79	13	2250	79	-	2857	2
16.0187	NS1/E2	680	LPCSFTTLPA	64	423	24000	9167	-	15	2
29.0039	Core	169	LPGCSFSI	92	500	200	932	620	6250	2
15.0219	Core	142	APLGGAARAL	71	9.5	-	-	-	12500	1
29.0031	NS5	2869	APTLWARM	79	13	-	4583	-	4348	1
15.0231	NS3	1512	RPSGMFDSSV	71	153	-	-	-	-	1
29.0085	NS5	2474	LPINALSNSL	57	220	18000	1170	-	11111	1
29.0037	NS5	2608	KPARLIVF	85	367	-	3235	-	16667	1
15.0237	NS4	1789	NPAIASLMAF	71	393	9000	5000	-	-	1
29.0118	NS5	2869	APTLWARMILM	79	423	-	-	-	3030	1
29.0042	NS4	1720	LPYIEQGM	85	423	-	1375	-	7692	1

C. Engineered analogs of B7 supermotif peptides.

					B7-supertype binding capacity (IC50 nM)					
Peptide	Molecule	1st Position	Sequence	Consv.	B*0702	B*3501	B*51	B*5301	B*5401	B7 XRN
1145.12	Core	169	LPGCSFSIF	92	28	90	100	114	6667	4
1292.24	Core	169	LPGCSFSII		37	4364	5.3	262	1056	3
1145.13	Core	169	FPGCSFSIF		19	1.6	132	3.2	6.7	5

* A dash indicates IC50 nM >30,000.

Table XXIX: HCV-derived A1- and A24-motif containing peptides

A. A1-motif peptides

Peptide	Molecule	Position	Sequence	Conserv.	HLA-A*0101 binding (IC50 nM)
13.0019	NS5	2922	LSAFSLHSY	79	31
1.0509	NS5	2921	GLSAFSLHSY	79	61
1069.62	NS3	1128	CTCGSSDLY	79	68
24.0093	NS5	2129	EVDGVRLHRY	100	167
13.0016	NS3	1241	KSTKVPAAY	85	1923
1.0125	NS3	1525	CYDAGCAWY	79	4032
24.0008	E1	206	DCSNSSIVY	85	16667
24.0094	NS5	2720	TNSKGQNCGY	100	-
24.0096	NS3	1240	GKSTKVPAAY	85	-
24.0100	NS3	1292	TGAPITYSTY	85	-
	NS3	1263	VAATLGFGAY	100	
	NS5	2639	VMGSSYGFQY	79	
	NS5	2640	MGSSYGFQY	79	

A dash indicates IC50 nM >25000

B. A24 -motif peptides

Peptide	Molecule	Position	Sequence	Conserv.	HLA-A*2402 binding (IC50 nM)
24.0092	NS4	1765	FWAKHMWNF	85	1.7
13.0075	NS4	1778	QYLAGLSTL	100	250
1073.18	NS1/E2	636	MYVGGVEHRL	92	444
13.0074	NS3	1297	TYSTYGKFL	85	522
13.0134	NS5	2647	QYSPGQRVEF	79	667
24.0091	NS4	1772	NFISGIQYL	100	706
13.0131	Core	135	GYIPLVGAPL	79	2105
24.0108	Core	173	SFSIFLLALL	100	2927
13.0132	NS3	1248	AYAAQGYKVL	79	13333
13.0133	NS4	1859	GYGAGVAGAL	85	-
1174.08	NS4	1769	HMWNFISGI	93	
	E1	317	RMAWDMMMNW	85	
	NS1/E2	635	RMYVGGVEHRL	93	
	NS3	1422	YYRGLDVSVI	100	
	NS3	1468	DFSLDPTFTI	100	
	NS3	1608	SWDQMWKCL	79	
	NS3	1664	TWVLVGGVL	85	
	NS4	1732	QFKQKALGL	85	
	NS4	1732	QFKQKALGLL	85	
	NS4	1765	FWAKHMWNFI	85	
	NS4	1919	QWMNRLIAF	100	
	NS5	2241	LWRQEMGGNI	85	
	NS5	2669	GFSYDTRCF	79	
	NS5	2875	RMILMTHFF	85	

A dash indicates IC50 nM >25000

Table XXX: Immunogenicity of A2-supertype cross-reactive binders

				Immunogenicity						
				Human[a]					Transgenic mice[b]	
Peptide	Sequence	Protein	Position	Barnaba; patients	Barnaba; contacts	Chisari	Pape	overall	Frequency	Response
1073.05	LLFNILGGWV	NS4	1812	1/6	7/17	2/21	0/6	10/50	6/6	6.4 (1.7)
1090.18	FLLLADARV	NS1/E2	728	2/6	7/17	1/21	0/6	10/50	5/6	9.5 (3.0)
1013.02	YLVAYQATV	NS4	1590	1/6	4/17	1/21	0/6	6/50	5/6	8.5 (3.7)
1090.22	RLIVFPDLGV	NS5	2578	2/6	5/17	0/21	0/6	7/50	0/6	-
1013.1002	DLMGYIPLV	Core	132	2/6	7/17	1/21	1/6	11/50	5/6	8.8 (2.6)
24.0073	WMNRLIAFA	NS4	1920	1/6	3/17	2/21	1/6	7/50	0/6	-
24.0075	VLVGGVLAA	NS4	1666	1/6	6/17	3/21	1/6	11/50	0/6	-
1174.08	HMWNFISGI	NS4	1769	3/6	3/17	2/21	0/6	8/50	6/6	6.4 (1.7)
1073.06	ILAGYGAGV	NS4	1851	2/6	3/17	0/21	0/6	5/50	3/6	54.7 (3.3)
1073.07	YLLPRRGPRL	CORE	35	2/6	5/17	7/21	1/6	17/50	4/6	59.1 (7.2)
24.0071	LLFLLLADA	NS1/E2	726	2/6	9/17	0/21	0/6	11/50	0/6	-
1.0119	YLVTRHADV	NS3	1131	6/6	10/17	0/21	1/6	17/50	0/6	-

a. Data shown represents the number of positive responses over the total number of patients or contacts examined.

b. Frequency represents the number of positive responses over the total number of mice examined. Response indicates the average magnitude (standard deviation) of the response in positive animals, measured in lytic units.

Table XXXI: Immunogenicity of A3-supertype cross-reactive binders

				Immunogenicity						
				Human[a]					Transgenic mice[b]	
Peptide	Sequence	Protein	Position	Barnaba patients	Barnaba; contacts	Chisari	Pape	overall	Frequency	Response
1.0952	KTSERSQPR	CORE	51	2/16	1/4	3/12	0/6	6/38	3/6	23.4 (1.3)
1073.11	RLGVRATRK	CORE	43	4/16	1/4	7/12	1/6	13/38	3/6	42.2 (1.2)
1.0955	QLFTFSPRR	ENV	290	1/16	0/4	6/12	1/6	8/38		
1073.13	RMYVGGVEHR	NS1/E2	632	5/16	1/4	4/12	1/6	11/38	2/6	2.8 (1.1)
1.0123	LIFCHSKKK	NS3	1396	6/16	1/4	4/12	2/6	13/38	3/6	4.4 (1.1)
1073.10	GVAGALVAFK	NS4	1863	3/16	0/4	6/12	2/6	11/38	6/6	56.5 (1.7)
24.0090	VAGALVAFK	NS4	1864	4/16	1/4	6/12	0/4	11/38	1/6	7.1
24.0086	TLGFGAYMSK	NS3	1262	6/16		2/12	2/5	10/33		

a. Data shown represents the number of positive responses over the total number of patients or contacts examined.

b. Frequency represents the number of positive responses over the total number of mice examined. Response indicates the average magnitude (standard deviation) of the response in positive animals, measured in lytic units.

Table XXXII. Candidate HCV-derived HTL epitopes

Selection criteria	Peptide	Sequence	Source	Conservancy Total	Conservancy Core
A. DR-supermotif conserved 15mers	1283.01	GQIVGGVYLLPRRGPR	HCV Core 28	93	93
	1283.02	VYLLPRRGPRLGVRA	HCV Core 34	93	93
	1283.03	GWLLSPRGSRPSWGPT	HCV Core 95	79	79
	1283.04	LGKVIDTLTCGFADL	HCV Core 119	79	86
	1283.05	IDTLTCGFADLMGYI	HCV Core 123	86	86
	1283.06	ADLMGYIPLVGAPLG	HCV Core 131	79	79
	1283.07	GVRVLEDGVNYATGN	HCV Core 154	86	86
	1283.08	GVNYATGNLPGCSFS	HCV Core 161	79	86
	1283.09	GCSFSIFLLALLSCL	HCV Core 171	86	100
	1283.10	GHRMAWDMMMNWSPT	HCV E1 315	86	86
	1283.11	CGPVYCFTPSPVVVG	HCV NS1/E2 506	93	93
	1283.12	VYCFTPSPVVVGTTD	HCV NS1/E2 509	93	93
	1283.13	GNWFGCTWMNSTGFT	HCV NS1/E2 550	79	86
	1283.14	FTTLPALSTGLIHLH	HCV NS1/E2 684	79	86
	1283.17	DLYLVTRHADVIPVR	HCV NS3 1134	79	79
	1283.18	RAAVCTRGVAKAVDF	HCV NS3 1186	79	79
	1283.20	AQGYKVLVLNPSVAA	HCV NS3 1251	79	100
	1283.21	GYKVLVLNPSVAATL	HCV NS3 1253	100	100
	1283.22	VLVLNPSVAATLGFG	HCV NS3 1256	100	100
	1283.23	GTVLDQAETAGARLV	HCV NS3 1335	86	86
	1283.24	GARLVVLATATPPGS	HCV NS3 1345	79	86
	1283.25	GRHLIFCHSKKKCDE	HCV NS3 1393	100	100
	1283.27	DSVIDCNTCVTQTVD	HCV NS3 1454	86	86
	1283.28	TVDFSLDPTFTIETT	HCV NS3 1466	79	100
	1283.30	FTGLTHIDAHFLSQT	HCV NS3 1567	93	93
	1283.31	YLVAYQATVCARAQA	HCV NS3 1591	79	93
	1283.32	KPTLHGPTPLLYRLG	HCV NS4 1620	79	79
	1283.33	LEVVTSTWVLVGGVL	HCV NS4 1658	86	86
	1283.34	TWVLVGGVLAALAAY	HCV NS4 1664	86	86
	1283.35	AEQFKQKALGLLQTA	HCV NS4 1730	86	86
	1283.40	PAILSPGALVVGVVCA	HCV NS4 1889	79	93
	1283.41	GALVVGVVCAAILRR	HCV NS4 1895	79	79
	1283.42	CAAILRRHVGPGEGA	HCV NS4 1903	79	79
	1283.43	AVQWMNRLIAFASRG	HCV NS4 1917	100	100
	1283.44	MNRLIAFASRGNHVS	HCV NS4 1921	86	100
	1283.48	ANLLWRQEMGGNITR	HCV NS5 2238	86	86
	1283.49	RQEMGGNITRVESEN	HCV NS5 2243	86	86
	1283.52	ARLIVFPDLGVRVCE	HCV NS5 2610	79	79
	1283.53	FPDLGVRVCEKMALY	HCV NS5 2615	79	100
	1283.54	GVRVCEKMALYDVVS	HCV NS5 2619	79	100
	1283.56	QPEYDLELITSCSSN	HCV NS5 2808	79	93
	1283.57	LELITSCSSNVSVAH	HCV NS5 2813	79	100
	1283.58	PTLWARMILMTHFFS	HCV NS5 2870	79	86
	1283.59	LHGLSAFSLHSYSPG	HCV NS5 2919	79	79
	1283.60	AFSLHSYSPGEINRV	HCV NS5 2924	79	79

Table XXXII. Candidate HCV-derived HTL epitopes

Selection criteria	Peptide	Sequence	Source	Conservancy Total	Conservancy Core
B. High algorithm conserved core	1283.15	VVLLFLLLADARVCS	HCV NS1/E2 724	29	100
	1283.16	SKGWRLLAPITAYAQ	HCV NS3 1025	29	79
	1283.19	PQTFQVAHLHAPTGS	HCV NS3 1225	43	85
	1283.26	DVVVVATDALMTGYT	HCV NS3 1436	43	79
	1283.29	WESVFTGLTHIDAHF	HCV NS3 1563	43	92
	1283.45	LTSMLTDPSHITAET	HCV NS5 2176	57	100
	1283.46	ASQLSAPSLKATCTT	HCV NS5 2208	50	79
	1283.47	DADLIEANLLWRQEM	HCV NS5 2232	50	85
	1283.50	SYTWTGALITPCAAE	HCV NS5 2456	64	79
	1283.51	TTIMAKNEVFCVQPE	HCV NS5 2589	64	85
	1283.55	GSSYGFQYSPGQRVE	HCV NS5 2641	71	79
	1283.61	ASCLRKLGVPPLRVW	HCV NS5 2939	50	85
C. Collaborator	F098.03	AAYAAQGYKVLVLNPSVAAT	HCV NS3 1242-1261	71	100
	F098.04	GYKVLVLNPSVAATLGFGAY	HCV NS3 1248-1267	100	
	F098.05	GYKVLVLNPSVAAT	HCV NS3 1248-1261	100	
	F134.01	RRPQDVKFPGGGQIVGGVY	HCV Core 17-35	86	
	F134.02	DVKFPGGGQIVGGVYLLPRR	HCV Core 21-40	86	
	F134.03	GYKVLVLNPSVAATLGFGAY	HCV NS3 1253-1272	100	
	F134.04	TLHGPTPLLYRLGAVQNEIT	HCV NS4 1622-1641		79
	F134.05	NFISGIQYLAGLSTLPGNPA	HCV NS4 1772-1791	100	
	F134.06	LLFNILGGWVAAQLAAPGAA	HCV NS4 1812-1831		86
	F134.07	GPGEGAVQWMNRLIAFASRG	HCV NS4 1912-1931	86	100
	F134.08	GEGAVQWMNRLIAFASRGNHV	HCV NS4 1914-1934	100	
	Pape 21	AIPLEVIKGGRHLIFCHSKR	HCV NS3 1379-1398	21	100
	Pape 22	GRHLIFCHSKRKCDELATKL	HCV NS3 1388-1407		100
	Pape 29	SVIDCNTCVTQTVDFSLDPT	HCV NS3 1450-1469	86	
D. DR3 motif	35.0102	GVRVLEDGVNYATGN	HCV 154	86	86
	35.0103	SAMYVGDLCGSVFLV	HCV 273	57	86
	35.0104	GHRMAWDMMMNWSPT	HCV 315	86	86
	35.0105	SDLYLVTRHADVIPV	HCV 1133	79	86
	35.0106	VVVVATDALMTGYTG	HCV 1437	42	86
	35.0107	TVDFSLDPTFTIETT	HCV 1466	79	100
	35.0108	DSSVLCECYDAGCAW	HCV 1518	71	93
	35.0109	GLPVCQDHLEFWESV	HCV 1552	42	86
	35.0110	GMQLAEQFKQKALGL	HCV 1726	57	86
	35.0111	PTHYVPESDAAARVT	HCV 1936	86	86
	35.0112	GSQLPCEPEPDVAVL	HCV 2162	64	86
	35.0113	LTSMLTDPSHITAET	HCV 2176	57	100
	35.0114	MPPLEGEPGDPDLSD	HCV 2401	79	100
	35.0115	QPEYDLELITSCSSN	HCV 2808	79	93
	1283.25	GRHLIFCHSKKKCDE	HCV NS3 1393-1407		

Table XXXIII. HLA-DR screening panels

Screening Panel	Antigen	Alleles	Representative Assay		Phenotypic Frequencies					
			Allele	Alias	Cauc.	Blk.	Jpn.	Chn.	Hisp.	Avg.
Primary	DR1	DRB1*0101-03	DRB1*0101	(DR1)	18.5	8.4	10.7	4.5	10.1	10.4
	DR4	DRB1*0401-12	DRB1*0401	(DR4w4)	23.6	6.1	40.4	21.9	29.8	24.4
	DR7	DRB1*0701-02	DRB1*0701	(DR7)	26.2	11.1	1.0	15.0	16.6	14.0
	Panel total				59.6	24.5	49.3	38.7	51.1	44.6
Secondary	DR2	DRB1*1501-03	DRB1*1501	(DR2w2 ß1)	19.9	14.8	30.9	22.0	15.0	20.5
	DR2	DRB5*0101	DRB5*0101	(DR2w2 ß2)	-	-	-	-	-	-
	DR9	DRB1*09011,09012	DRB1*0901	(DR9)	3.6	4.7	24.5	19.9	6.7	11.9
	DR13	DRB1*1301-06	DRB1*1302	(DR6w19)	21.7	16.5	14.6	12.2	10.5	15.1
	Panel total				42.0	33.9	61.0	48.9	30.5	43.2
Tertiary	DR4	DRB1*0405	DRB1*0405	(DR4w15)	-	-	-	-	-	-
	DR8	DRB1*0801-5	DRB1*0802	(DR8w2)	5.5	10.9	25.0	10.7	23.3	15.1
	DR11	DRB1*1101-05	DRB1*1101	(DR5w11)	17.0	18.0	4.9	19.4	18.1	15.5
	Panel total				22.0	27.8	29.2	29.0	39.0	29.4
Quarternary	DR3	DRB1*0301-2	DRB1*0301	(DR3w17)	17.7	19.5	0.4	7.3	14.4	11.9
	DR12	DRB1*1201-02	DRB1*1201	(DR5w12)	2.8	5.5	13.1	17.6	5.7	8.9
	Panel total				20.2	24.4	13.5	24.2	19.7	20.4

Table XXXIV. HLA-DR binding capacity of target derived peptides: DR-supermotif and algorithm positive peptides.

Peptide	Sequence	Source	Binding capacity (IC50 nM)											DR alleles
			DR1	DR2w2β1	DR2w2β2	DR4w4	DR4w15	DR5w11	DR6w19	DR7	DR8w2	DR9	IAb	bound
	AAYAAQGYKVLVLNPSVAATLGFGAY	HCV NS3 1242-1267												
1283.21	GYKVLVLNPSVAATL	HCV NS3 1253	4.5	350		5.2	567	143	5.1	89	288	54	175	9
1283.20	AQGYKVLVLNPSVAA	HCV NS3 1251	6.0	650		7.9	224	74	5.9	833	175	375	298	9
F98.03	AAYAAQGYKVLVLNPSVAAT	HCV NS3 1242	2.9	48	483	18	1234	103	11	96	60	240		9
F98.05	GYKVLVLNPSVAAT	HCV NS3 1248-1261	1.4	39	3695	7.8	141	75	3.5	126	21	266		9
F98.04	GYKVLVLNPSVAATLGFGAY	HCV NS3 1248-1267	3.5	42	8154	9.7	1500	240	4.1	23	80	20		8
	GEGAVQWMNRLIAFASRGNHVS	HCV NS4 1914-1935												
1283.44	MNRLIAFASRGNHVS	HCV NS4 1921	66	4.8	1538	6329	585	45	7.3	227	102	313	147	8
F134.08	GEGAVQWMNRLIAFASRGNHV	HCV NS4 1914	3.2		182	361		345		221	158	6818		6
1283.16	SKGWRLLAPITAYAQ	HCV NS3 1025	0.36	125	23	24	152	4.8		962	54	1190	384	8
1283.55	GSSYGFQYSPGQRVE	HCV NS5 2641	11		667	417	745	20000	19	156		68	571	7
1283.61	ASCLRKLGVPPLRVW	HCV NS5 2939	5.0	16	217	6250	78	645	2500	862	671	8621	-	7
F134.05	NFISGIQYLAGLSTLPGNPA	HCV NS4 1772	10		606	84		29			70	441		6

Shading indicates IC50 >1 μM.
A dash (-) indicates IC50 >20 μM.

Table XXXV. HLA-DR binding capacity of 3 DR3 motif-containing peptides

Peptide	Sequence	Source	DR3 binding (IC50 nM)
35.0106	VVVVATDALMTGYTG	HCV 1437	427
35.0107	TVDFSLDPTFTIETT	HCV 1466	235
1283.25	GRHLIFCHSKKKCDE	HCV NS3 1393	ND

10073345 .072202

Table XXXVIa: HCV-derived CTL epitope candidates

Peptide	Molecule	1st Position	Sequence	Consv.	Selection criteria
1073.05	NS4	1812	LLFNILGGWV	85	A2-supertype
1090.18	NS1/E2	728	FLLLADARV	92	A2-supertype
1013.02	NS4	1590	YLVAYQATV	85	A2-supertype
1090.22	NS5	2611	RLIVFPDLGV	79	A2-supertype
1013.1002	CORE	132	DLMGYIPLV	79	A2-supertype
24.0073	NS4	1920	WMNRLIAFA	100	A2-supertype
24.0075	NS4	1666	VLVGGVLAA	85	A2-supertype
1174.08	NS4	1769	HMWNFISGI	92	A2-supertype
1073.06	NS4	1851	ILAGYGAGV	79	A2-supertype
1073.07	CORE	35	YLLPRRGPRL	92	A2-supertype
24.0071	NS1/E2	726	LLFLLLADA	100	A2-supertype
1.0119	LORF	1131	YLVTRHADV	85	A2-supertype
1.0952	CORE	51	KTSERSQPR	92	A3-supertype
1073.11	CORE	43	RLGVRATRK	79	A3-supertype
1.0955	ENV1	290	QLFTFSPRR	79	A3-supertype
1073.13	NS1/E2	632	RMYVGGVEHR	100	A3-supertype
1.0123	NS3	1396	LIFCHSKKK	100	A3-supertype
1073.10	NS4	1863	GVAGALVAFK	85	A3-supertype
24.0090	NS4	1864	VAGALVAFK	85	A3-supertype
24.0086	NS3	1262	TLGFGAYMSK	85	A3-supertype
F104.01	NS5	3003	VGIYLLPNR	79	A31
1145.12	Core	169	LPGCSFSIF	92	B7-supertype
29.0035	NS3	1378	IPFYGKAI	92	B7-supertype
13.0019	NS5	2922	LSAFSLHSY	79	A1
1069.62	NS3	1128	CTCGSSDLY	79	A1
24.0092	NS4	1765	FWAKHMWNF	85	A24

Table XXXVIb: HCV-derived HTL epitope candidates

Region	Peptide	Motif[1]	Sequence
HCV NS3 1025-1039	1283.16	DR	SKGWRLLAPITAYAQ
HCV NS3 1242-1267	F98.03	DR	AAYAAQGYKVLVLNPSVAAT
HCV NS3 1393-1407	1283.25	DR3	GRHLIFCHSKKKCDE
HCV NS3 1437-1451	35.0106	DR3	VVVVATDALMTGYTG
HCV NS3 1466-1480	35.0107	DR3	TVDFSLDPTFTIETT
HCV NS4 1772-1790	F134.05	DR	NFISGIQYLAGLSTLPGNPA
HCV NS4 1914-1935	F134.08	DR	GEGAVQWMNRLIAFASRGNHV
HCV NS5 2641-2655	1283.55	DR	GSSYGFQYSPGQRVE
HCV NS5 2939-2953	1283.61	DR	ASCLRKLGVPPLRVW

1. Peptides identified on the basis of either the DR P1-P6 supermotif or by use of the DR1-4-7 algorithms are indicated by 'DR'. Peptides identified using the DR3 motif are indicated by 'DR3'.

Table XXXVII. Estimated population coverage by a panel of HCV derived HTL epitopes

Antigen	Alleles	Representative assay	No. of epitopes[2]	Population coverage (phenotypic frequency) Cauc.	Blk.	Jpn.	Chn.	Hisp.	Avg.
DR1	DRB1*0101-03	DR1	6	18.5	8.4	10.7	4.5	10.1	10.4
DR2	DRB1*1501-03	DR2w2 ß1	3	19.9	14.8	30.9	22.0	15.0	20.5
DR2	DRB5*0101	DR2w2 ß2	6	-	-	-	-	-	-
DR3	DRB1*0301-2	DR3	2	17.7	19.5	0.40	7.3	14.4	11.9
DR4	DRB1*0401-12	DR4w4	5	23.6	6.1	40.4	21.9	29.8	24.4
DR4	DRB1*0401-12	DR4w15	3	-	-	-	-	-	-
DR7	DRB1*0701-02	DR7	5	26.2	11.1	1.0	15.0	16.6	14.0
DR8	DRB1*0801-5	DR8w2	5	5.5	10.9	25.0	10.7	23.3	15.1
DR9	DRB1*09011,09012	DR9	3	3.6	4.7	24.5	19.9	6.7	11.9
DR11	DRB1*1101-05	DR5w11	5	17.0	18.0	4.9	19.4	18.1	15.5
DR13	DRB1*1301-06	DR6w19	2	21.7	16.5	14.6	12.2	10.5	15.1
Total[1]				98.5	95.1	97.1	91.3	94.3	95.1

1. Total population coverage has been adjusted to acount for the presence of DRX in many ethnic populations. It has been assumed that the range of specificities represented by DRX alleles will mirror those of previously characterized HLA-DR alleles. The proportion of DRX incorporated under each motif is representative of the frequency of the motif in the remainder of the population. Total coverage has not been adjusted to account for unknown gene types.
2. Number of epitopes represents a minimal estimate, considering only the epitopes shown in Table 6. Additional alleles possibly bound by nested epitopes have not been accounted.

TABLE Ia

SUPERMOTIFS	POSITION	POSITION	POSITION
	2 (Primary Anchor)	3 (Primary Anchor)	C Terminus (Primary Anchor)
A1	**T, I,** *L, V, M, S*		**F, W, Y**
A2	**V,** *Q, A, T*		**I, V,** *L, M, A, T*
A3	**V, S, M, A,** *T, L, I*		**R, K**
A24	**Y, F,** *W, I, V, L, M, T*		**F, I,** *Y, W, L, M*
B7	**P**		**V, I, L, F,** *M, W, Y, A*
B27	**R, H, K**		**F, Y, L,** *W, M, I, V, A*
B58	**A, T, S**		**F, W, Y,** *L, I, V, M, A*
B62	**Q, L,** *I, V, M, P*		**F, W, Y,** *M, I, V, L, A*
MOTIFS			
A1	**T, S, M**		**Y**
A1		**D, E,** *A, S*	**Y**
A2.1	*V, Q, A, T**		**V,** *L, I, M, A, T*
A3.2	**L, M, V, I, S, A, T, F,** *C, G, D*		**K, Y, R,** *H, F, A*
A11	**V, T, M, L, I, S, A, G, N,** *C, D, F*		**K,** *R, H, Y*
A24	**Y ,F, W**		**F, L, I, W**

*If 2 is V, or Q, the C-term is not L

Bolded residues are preferred, italicized residues are less preferred: A peptide is considered motif-bearing if it has primary anchors at each primary anchor position for a motif or supermotif as specified in the above table.

SF 1116265 v1